



University of Pennsylvania
ScholarlyCommons

Finance Papers

Wharton Faculty Research

12-2016

Blind Queues: The Impact of Consumer Beliefs on Revenues and Congestion

Shiliang Cui

Senthil K. Veeraraghavan
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/fnce_papers



Part of the [Finance and Financial Management Commons](#)

Recommended Citation

Cui, S., & Veeraraghavan, S. K. (2016). Blind Queues: The Impact of Consumer Beliefs on Revenues and Congestion. *Management Science*, 62 (12), 3656-3672. <http://dx.doi.org/10.1287/mnsc.2015.2320>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/fnce_papers/125
For more information, please contact repository@pobox.upenn.edu.

Blind Queues: The Impact of Consumer Beliefs on Revenues and Congestion

Abstract

In many service settings, customers have to join the queue without being fully aware of the parameters of the service provider (e.g., customers at checkout counters may not know the true service rate before joining). In such “blind queues,” customers make their joining/balking decisions based on limited information about the service provider’s operational parameters (from past service experiences, reviews, etc.) and queue lengths. We analyze a firm serving customers making decisions under arbitrary beliefs about the service parameters in an observable queue for a service with a known price. By proposing an ordering for the balking threshold distributions in the customer population, we are able to compare the effects of customer beliefs on the queue. We show that, although revealing the service information to customers improves revenues under certain conditions, it may destroy consumer welfare or social welfare. Given a market size, consumer welfare can be significantly reduced when a fast server announces its true service parameter. When revenue is higher under some beliefs, one would expect the congestion to also be higher because more customers join, but we show that congestion may not necessarily increase.

Keywords

consumer decisions, lack of information, service revelation, consumer beliefs, queueing

Disciplines

Finance and Financial Management

Blind Queues: The Impact of Consumer Beliefs on Revenues and Congestion

Shiliang Cui* Senthil Veeraraghavan^{†‡}

April 2015

In many service settings, customers have to join the queue without being fully aware of the parameters of the service provider (for e.g., customers at check-out counters may not know the true service rate prior to joining). In such “blind queues”, customers make their joining/balking decisions based on the limited information about the service provider’s operational parameters (from past service experiences, reviews, etc.), and queue lengths. We analyze a firm serving customers making decisions under arbitrary beliefs about the service parameters in an observable queue for a service with known price.

We propose an ordering for the balking threshold distributions in customer population, which allows for comparing the effects of customer beliefs on the queue. We show that while revealing the service information to customers improves revenues under certain conditions, it may however destroy consumer welfare or social welfare. Given a market size, the consumer welfare can be significantly reduced when a fast server announces its true service parameter. When revenue is higher under some beliefs, one would expect the congestion to also be higher because more customers join, but we show the congestion may not necessarily increase.

Key words: Customer Decisions, Lack of Information, Service Revelation, Customer Beliefs, Queueing.

1. Introduction

Almost all the literature on queues assumes that the service parameters are common knowledge and fully known to customers when making their decisions. In reality, it is likely that only the service firm knows its capacity and the customers may not be fully informed of the service capacity. It is even possible that customers could be systematically misinformed about a firm’s service capacity. Hence, it is important to understand if the firm is motivated to reveal its private information on service rate to the customers, and if the firm does reveal the information, whether the information would increase consumer welfare or firm revenues. To be sure, there are papers that have focused

* McDonough School of Business, Georgetown University, Washington, DC, 20057. EML: shiliang.cui@georgetown.edu

[†] The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104. EML: senthilv@wharton.upenn.edu

[‡] Our sincere thanks to suggestions from Gad Allon, Gerard Cachon, Teck Ho, Noah Gans, Costis Maglaras, Sergei Savin, Xuanming Su, Assaf Zeevi, an associate editor and two anonymous reviewers that led to improvements in the manuscript. Research funded in part by Fishman-Davidson Center.

on firms announcing (real-time) delay information in terms of queue and waiting time information to its customers. However, those models also generally assume that the firm’s service parameters are known to the customers.

We expect that customers that have had limited past interactions with the service provider will not be able to accurately predict its true service rate. For instance, a customer might have visited a restaurant or an amusement park only once or twice, and it is conceivable that her best estimate of the service capacity will be based on the service times she had experienced (in the absence of other inputs). In some cases, customers might augment their information using feedback from external acquaintances, but even such information is likely to be a smaller sample than what is needed to know the full service distribution (which is often assumed to be known accurately in the literature). In line with many real-life services, but in contrast to the existing literature, we allow for the customers to *not know* the service parameter accurately, unless they are informed about it by the firm. We term such queues as *blind queues*.

Our approach to the analysis is general, i.e., individual customers can have arbitrarily different beliefs about the service rate. It might be possible that the population is correct on average but individual customers may be idiosyncratically misinformed. We also consider the possibility that the population as a whole is mis-informed systematically. In our model, queue joining (or balking) decision is irrevocable. Thus, we do not consider reneging behavior, that may result from in-queue learning. However, it is possible that the in-queue learning accrued during past experiences or out-of-queue learning form the customer beliefs in our model that drive joining or balking.

In observable queues, when customers arrive with different beliefs about the true service rate, they exhibit different balking behaviors due to their beliefs. For instance, a customer may join the queue when he is better off with a balking decision (if he overestimates the true service rate), and conversely, a misinformed customer may balk from the queue when he should not.

Note that customers’ internal beliefs about the service rate are not observable; the server only observes customers’ joining/balking decisions. Hence, we work primarily with the observable balking threshold distribution that results from the original service-rate belief distribution, but address the relationship between the distributions at the end of our main analysis. There is some recent empirical evidence (see Lu et al. (2013)) that uses queueing data from a Deli), supporting the approach that customers in observable queues may rely primarily on the length of the queue to make their purchasing/joining decisions. Thus, by understanding the impact of balking thresholds on system performance, our results could be further implemented to models where service values and waiting costs are heterogeneous.

1.1. Related Literature

The literature on queueing models with strategic customers dates back to the seminal paper of Naor (1969), who studies a single-server system with an observable queue. In Naor's model, homogeneous customers (who *know* the service parameters) observe the queue length upon arrival before making a decision to join the system. Because of homogeneity, customers have identical balking thresholds.

In our paper, customers are not aware of the true service parameters. We allow them to have arbitrarily distributed heterogeneous beliefs over the service rate. Thus, our work is also closely related to the classical queueing papers with heterogeneous customers (with full information), in addition to those papers that examine the effect of delay announcements. Following Naor (1969), queues with heterogeneous service values and time costs have been studied, as seen in the comprehensive review by Hassin and Haviv (2003).

There is a large volume of literature that examines the provision of fixed or variable *delay* information (i.e. queue length or real-time waiting time, etc.) to arriving customers. In the context of call-centers, there are several papers that study the provision of current delay information. For instance, see Armony and Maglaras (2004a,b) and Jouini et al. (2011). We refer the reader to an excellent review by Akşin et al. (2007) of the call center literature, on the role of delay information on customers' balking behavior. Nevertheless, the service capacity and arrival information is often assumed to be known to all customers in these papers.

Hassin (1986) considers a revenue-maximizing server who may hide queue lengths to improve revenue. Whitt (1999) shows that customers are more likely to be blocked in a system where the delay information is not provided to a system where it is provided. Guo and Zipkin (2007) studies an $M/M/1$ queue extension with three modes of information: no information, partial information (the queue length) and full information (the exact waiting time). Economou and Kanta (2008) and Guo and Zipkin (2009) study models where some partitioned queue information (such as range of queue-lengths) is available to customers to make their decisions. However, in all the aforementioned papers (including the no-information, unobservable queues first studied in Edelson and Hildebrand (1975)), the customers are aware of the service rate parameters.

Thus, there are very few papers that treat service information as a firm's private information about which customers are either not informed or have incorrect beliefs. Besbes et al. (2011) and Debo and Veeraraghavan (2014) analyze customers' equilibrium joining behaviors in queues with limited information on service rate. In Hassin (2007), the true service rate is either fast or slow. While the probabilities of the service rate being fast or slow is known to customers, the server can choose to reveal or not to reveal the realized rate. In Guo et al. (2011), partial distributional

information is conveyed to the customers, who then employ the max-entropy distribution in deciding whether to join or balk from the queues. In these papers, individual customers have correct distributional information over the service rate.

In contrast, we do not impose any such condition, i.e., individual customers could have incorrect information, and we study the impact of announcing service information in such cases. To the best of our knowledge, there does not exist any prior literature on customer decisions when they do not have *any* information about the service parameters.

Finally, our approach complements the perspective in Besbes and Maglaras (2009) and Haviv and Randhawa (2014), where the service firm does not fully know the demand (volume) information. Instead, we study a system where customers do not know a firm's service information. We focus only on the decision whether the firm should reveal the unknown information to the customers.

The main theoretical contributions of our paper can be summarized as follows:

1. Customers may have arbitrary balking thresholds due to their beliefs and decision frameworks. We characterize those balking threshold distributions under which the firm will gain revenues from revealing its service information.
2. We propose an order for customer balking threshold distributions, and show as they become less dispersed in the population, the firm improves its revenues. When the customer population systematically underestimates the service capacity, the resulting balking threshold distribution tends to be pessimistically biased, and the service provider should always reveal its service rate information in that case.
3. When revenue is higher under some balking threshold distributions, one would expect the congestion to also be higher because more customers join, but we show that this is not always true by providing counter-examples.
4. The welfare effects of information revealing are mixed. System congestion (both queue lengths and wait times) can increase with information revelation. Individual consumer welfare thus can worsen with more information, especially when a firm with high service rate releases its information. Social welfare may also fall with more information, because the improvement in revenues may be insufficient to overcome the consumer welfare loss.

The rest of the paper is structured as follows. Section 2 introduces the model and characterizes the system performance in terms of balking threshold distributions. In Section 3, we analyze customer populations with different distributions. In Section 4, we investigate the impact of the revelation of service information on revenues, system congestion, and welfare. Section 5, concludes the paper and addresses some future research directions. All proofs are deferred to the appendix.

2. Model

We focus on a single-server queueing system. (All results extend to a multi-server system.) Customers arrive to the queue according to a Poisson process at rate $\lambda > 0$ per unit time. The service time at the server is exponentially distributed with service rate $\mu > \lambda$. Let $\rho \triangleq \lambda/\mu < 1$ denote the traffic intensity. Arriving customers line up at the server if the server is busy, and the queue discipline is first-come first-served (FCFS). Every arriving customer observes the number of the customers already waiting in the system before making an irrevocable join or balk decision (i.e., there is no reneging). On joining, all customers incur a linear waiting cost of c per unit time when they wait. The server provides a service of value v . Thus, all customers are homogeneous in their service valuation and in their waiting costs. The firm charges an *exogenous* price p for its service.

Upon arrival, the customers decide whether to join the queue based on the net value they expect to receive from the service (i.e., $v - p$) and their expected waiting cost. Suppose the customers knew the true service rate μ . Then, a customer arriving when there are n customers already in the system, would join the queue when $v - p - (n + 1)c/\mu \geq 0$, or balk otherwise. This is the model described in Naor (1969).

Model of Customer (mis-)Information: In contrast to the existing literature, we relax the assumption that customers are aware of the service time distribution or the service rate. We posit that customers quite often will not have complete information on the true distribution. For instance, customers with limited or idiosyncratic past interactions with the server, may have widely varying service rate beliefs.

In this paper, we use the superscript \sim (tilde) to describe customer beliefs and the resulting distributions. Customers may have heterogeneous beliefs on the service rate, and their beliefs could differ arbitrarily from the true service rate μ . We denote customer service rate beliefs by $\tilde{\mu} \in (0, M)$ for some large M , with some cumulative distribution function (cdf) $G_{\tilde{\mu}}$ across the entire population. Note that every customer has a deterministic belief. The beliefs form a random distribution because customers with different beliefs arrive to the system randomly.

Now, we discuss *bias* using beliefs on the true service rate μ as an example. If the mean of the random variable, $\tilde{\mu}$, is equal to μ , i.e., the service rate belief of the population is “correct” on average, we describe the service rate beliefs of the population as being *consistent*. If the mean of $\tilde{\mu}$ is not μ , we address the service rate beliefs as being *biased*. Specifically, if the population mean is greater (less) than the true μ , the service rate beliefs are *optimistic* (*pessimistic*), i.e., the population is on average optimistic (pessimistic) on the speed of the server.

Upon arrival, a customer with belief $\tilde{\mu}$ who observes n customers currently waiting in the system (including the person who is under the service, if any) makes the following decision:

$$\begin{cases} v - p \geq \frac{(n+1)c}{\tilde{\mu}} : & \text{The customer joins the queue;} \\ \text{otherwise:} & \text{The customer balks from the queue.} \end{cases}$$

Throughout the paper, we will assume $v - p \geq \frac{c}{\tilde{\mu}}$ a.s. to eliminate trivial outcomes and ensure that customers will join an empty queue.

Balking Threshold Distribution: Define $\tilde{N} \triangleq \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor$, i.e., \tilde{N} is an integer such that $\tilde{N} \leq \frac{\tilde{\mu}(v-p)}{c} < \tilde{N} + 1$. Intuitively, \tilde{N} describes the *balking threshold* for a customer with service rate belief $\tilde{\mu}$: the customer who arrives to see n customers waiting in the system will join if $n + 1 \leq \tilde{N}$ and balk otherwise. Each customer has a deterministic balking threshold as a result of her (internally-held) service rate belief, and will make a deterministic decision upon seeing the queue length.

Let $F_{\tilde{N}}$ be the cdf that characterizes the random variable \tilde{N} . The balking thresholds are random because customers appear at random at the queue. Note that customers' beliefs on the service rate are drawn from some distribution $G_{\tilde{\mu}}$ which can be continuous, whereas the balking thresholds are drawn from a discrete distribution $F_{\tilde{N}}(n) = \Pr[\tilde{N} \leq n]$. Since $v - p \geq \frac{c}{\tilde{\mu}}$, we have $\tilde{N} \in \{1, 2, \dots\}$. In essence, we translate the customer beliefs on service rate to actions dictated by discrete balking thresholds. For notational convenience, we suppress the subscript \tilde{N} in $F_{\tilde{N}}$ and denote $F_{\tilde{N}}$ simply as F wherever unambiguous.

Note that the server only observes customers' joining/balking decisions, but not their service rate beliefs. Hence, we will directly analyze the balking threshold distribution \tilde{N} . Our terminology on biases in service rate beliefs (pessimism, optimism and consistency) also appropriately applies to balking threshold distributions. However, due to the floor function in the mapping from $\tilde{\mu}$ to \tilde{N} , the biases in service rate beliefs mostly follow the same direction as the biases in the corresponding balking threshold distribution, but not always. We will discuss the sufficient conditions in §4.

System Evolution under Balking Threshold Distribution: We have a population comprising of customers who are heterogeneous in their joining behavior due to the varying individual balking thresholds. Since $\tilde{N} \in \{1, 2, \dots\}$, we have a queuing system with state-dependent arrivals - a system whose buffer size equal to the maximum balking threshold. In contrast, note that when customers fully know μ , we get the classical $M/M/1/N$ system with state-independent arrivals that emerges in Naor (1969) where $N \triangleq \lfloor \frac{\mu(v-p)}{c} \rfloor$. Note that the system with heterogeneity in \tilde{N} which emerges from $\tilde{\mu}$ is similar in system evolution to a system with heterogeneity in v or c (see Hassin and Haviv (2003), pp. 34–36).

Let the state of system be denoted by i where i is the number of customers in the system (including the customer at the server). Since $\lambda < \mu$, this queueing system is recurrent, and long-run steady state probabilities exist. Let π_i denote the long-run probability that the system is in state i . Now consider state i : among all arrivals, only those customers who have the balking threshold greater than or equal to $i + 1$ will join the queue. Thus, the effective joining probability at state i is given by $\Pr[\tilde{N} \geq i + 1] = \Pr[\tilde{N} > i] = \bar{F}(i)$ (by letting $\bar{F}_{\tilde{N}}(\cdot) = 1 - F_{\tilde{N}}(\cdot)$). The effective arrival rate at any state i is $\lambda \bar{F}(i)$.

From the steady state rate balance equations, we have $\pi_{i+1} = \rho \bar{F}(i) \pi_i$ for $i \in \{0, 1, 2, \dots\}$, which gives $\pi_i = \rho^i \pi_0 \prod_{n=0}^{i-1} \bar{F}(n)$ for $i \in \{1, 2, 3, \dots\}$. Since $\rho < 1$, it follows from $\sum_{i=0}^{\infty} \pi_i = 1$ that

$$\pi_0 = 1 \left/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right. . \quad (1)$$

The average number of customers in the system, denoted by L , is given by

$$L = \sum_{i=0}^{\infty} i \pi_i = \sum_{i=1}^{\infty} i \pi_i = \pi_0 \sum_{i=1}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n) = \sum_{i=1}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \left/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right. . \quad (2)$$

By convention, we set all the empty products to 1. For instance, $\prod_{n=0}^{-1} \bar{F}(n) = 1$. Then,

$$L = \sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \left/ \left(\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right. . \quad (3)$$

The long-run revenue rate at the server, denoted by R , is given by $p\mu(1 - \pi_0)$, e.g., see Larsen (1998). It follows that the long-run effective arrival rate at the system, denoted by λ_{eff} , is:

$$\lambda_{eff} = \mu(1 - \pi_0) < \lambda. \quad (4)$$

Finally, let W denote the average time a customer spends in the system, i.e, her waiting time in the queue plus her service time. By Little's Law,

$$W = \frac{L}{\lambda_{eff}} = \frac{\pi_0 \sum_{i=1}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n)}{\mu(1 - \pi_0)} \quad (\text{from conditions (2) and (4)}) = \frac{1}{\mu} \frac{\pi_0}{1 - \pi_0} \sum_{i=1}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n). \quad (5)$$

Now recall from (1) that $\pi_0 = 1 \left/ \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right.$, hence, $\frac{\pi_0}{1 - \pi_0} = 1 \left/ \left(\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right.$. Plugging in (5), we have

$$W = \sum_{i=1}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \left/ \left(\mu \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}(n) \right) \right. . \quad (6)$$

Thus, as long as we can characterize the balking threshold distribution \tilde{N} , we can derive the performance measures for the queueing system through its cdf F . This allows us to compare any two systems with populations that differ in their balking threshold distributions. To this end, in the next section, we set up a sequence of systems with balking threshold distributions that are stochastically ordered in some sense.

3. Balking Thresholds under the Lack of Service Information

When μ is fully known to customers, the balking threshold distribution is a one-point distribution (i.e., all customers have the same balking threshold). In contrast, the balking thresholds are distributed arbitrarily when customers are not informed. Recall that when there is optimism (pessimism) bias, the average balking threshold is higher (lower) than the true threshold. In §3.1, we consider balking threshold belief distributions that have bias. In §3.2, we compare populations whose average balking threshold is accurate, but there is arbitrary variability on the individual thresholds. Our analysis assists us in pinning down the performance differences among queueing systems that differ in customer beliefs.

3.1. Population with Biased Balking Threshold Distribution

We first compare systems under balking threshold distributions \tilde{N} and \tilde{N}' that differ in their mean.

DEFINITION 1. *First Order Stochastic Dominance (FOSD):* (Quirk and Saposnik 1962) Let \mathcal{F} and \mathcal{G} be the cdf's of random variables X and Y . X is said to be smaller than Y with respect to the first-order stochastic order (written $X \leq_{st} Y$) if $\mathcal{F}(t) \geq \mathcal{G}(t)$ for all real t , or equivalently, if $\bar{\mathcal{F}}(t) \leq \bar{\mathcal{G}}(t)$ for all real t .

FOSD relation is also termed as *usual stochastic order* by Müller and Stoyan (2002), and frequently called *the stochastic order*. It is well known that variables ordered by FOSD have different means (e.g., see Theorem 1.2.9/(a) in Müller and Stoyan (2002)). Through FOSD relation, we can compare two balking threshold distributions with respect to their 'biases'. Essentially, pessimistic balking threshold distributions are stochastically dominated by more optimistic distributions.

We use $\lambda_{eff, \tilde{N}}, R_{\tilde{N}}, L_{\tilde{N}}, W_{\tilde{N}}$ to denote the long-run effective arrival (rate), the long-run revenue (rate) at the firm, the average number of customers and the average time a customer spends in the system when the balking threshold distribution of the population is characterized by \tilde{N} . The following Theorem 1 compares the performance metrics of two systems under balking threshold distributions ordered by FOSD.

THEOREM 1. *If $\tilde{N} \leq_{st} \tilde{N}'$, then (i) $\lambda_{eff, \tilde{N}} \leq \lambda_{eff, \tilde{N}'}$ ($R_{\tilde{N}} \leq R_{\tilde{N}'}$); (ii) $L_{\tilde{N}} \leq L_{\tilde{N}'}$; (iii) $W_{\tilde{N}} \leq W_{\tilde{N}'}$.*

Proof of Theorem 1 follows directly by comparing the stochastics of two queues, and by taking expectations of the distribution of the corresponding performance metrics (e.g., see Theorem 1 in Bhaskaran (1986)). Also see Berger and Whitt (1992) which compares two queues each with a single balking threshold. In our model, there are multiple balking thresholds in the same queue depending on the individual customer beliefs.

Theorem 1 states that if customers become more patient (higher balking thresholds), the firm serves more customers and receives higher revenue, at the same time incurring a higher system congestion. Note that the results are distribution-free, i.e., the performance metrics of queues can be ordered for any balking threshold distribution, as long as the underlying distributions can be (first-order) stochastically ordered. Thus, the ordering includes the distribution in which customers all share some common randomly drawn belief (and the corresponding balking threshold). Also, note that the ordering of performance metrics is invariant to the true service rate of the system.

3.2. Population with Consistent Balking Threshold Distribution

In this section, we consider mean-preserving spreads to examine balking threshold distributions that have the same mean as the “true” balking threshold, but differ in how the individual balking thresholds are distributed. If all customers knew the service rate exactly, then every customer uses the same “true” balking threshold N , which is a special consistent distribution. We use the notion of Single Mean Preserving Spread to order balking threshold distributions that are consistent.

DEFINITION 2. *Single Mean Preserving Spread (SMPS):* (Rothschild and Stiglitz 1970) Let \mathcal{F} and \mathcal{G} be the cdf’s of two discrete random variables X and Y whose common support is a sequence of real numbers $a_1 < a_2 < \dots < a_n$. Suppose the probability mass functions f and g describe X and Y completely: $\Pr(X = a_i) = f_i$ and $\Pr(Y = a_i) = g_i$ where $\sum_{i=1}^n f_i = \sum_{i=1}^n g_i = 1$. Suppose $f_i = g_i$ for all but four i , say i_1, i_2, i_3 and i_4 where $i_k < i_{k+1}$. Define $\gamma_{i_k} = g_{i_k} - f_{i_k}$. Then we say that Y differs from X by a single Mean Preserving Spread (written $\mathcal{F} \leq_{SMPS} \mathcal{G}$) if $\gamma_{i_1} = -\gamma_{i_2} \geq 0$, $\gamma_{i_4} = -\gamma_{i_3} \geq 0$ and $\sum_{k=1}^4 a_{i_k} \gamma_{i_k} = 0$.

The notion of mean preserving spread (MPS) is often employed to model risk ordering of two random variables that may have the same mean but different variability. If two distributions \mathcal{F} and \mathcal{G} describe the returns of two risky investments, and $\mathcal{F} \leq_{MPS} \mathcal{G}$, then distribution \mathcal{F} is considered less risky. SMPS in Definition 2 is a stricter condition than MPS: $\mathcal{F} \leq_{SMPS} \mathcal{G} \Rightarrow \mathcal{F} \leq_{MPS} \mathcal{G}$.

Consider consistent balking threshold distribution \tilde{N} , i.e., $\mathbb{E}[\tilde{N}]$ equals the balking threshold N (when the service parameters are fully known to the customers). We seek to compare the performance metrics under \tilde{N} to the metrics if the true parameters of the system were known. To that end, we create a sequence of random variables ordered by SMPS that begin at an initial

distribution. Using a fairly general, but intuitive construction technique, we will show that the sequence (generated using our construction) will terminate at a specific “final” distribution within finite number of steps, regardless of the initial distribution. We then characterize an ordering of the performance metrics for the entire sequence. This not only allows us to compare the performance under the initial distribution to the canonical system with fully informed customers, but also facilitates a comparison between any two arbitrary (consistent) balking threshold distributions.

Let our initial balking thresholds be characterized by some random variable \tilde{N}_0 . In Construction 1, we will create a sequence of random variables $\{\tilde{N}_K\}_{K \geq 0}$ (the K -th term in the sequence is distributed with the cdf F_K), and discuss the properties of the sequence. The cdf F_K has support over some finite sequence of natural numbers $a_{K_1} < a_{K_2} < \dots < a_{K_n}$. We denote f_K its probability mass function (pmf) such that $f_K(a_{K_i}) > 0$ for $i \in \{1, 2, \dots, n\}$ and $\sum_{i=1}^n f_K(a_{K_i}) = 1$.

Consider the transformation of \tilde{N}_K to \tilde{N}_{K+1} in the following Construction 1. The succeeding random variable in the sequence, \tilde{N}_{K+1} , is constructed from the preceding random variable \tilde{N}_K by taking an equal probability mass from both ends of the distribution F_K and adding those weights towards the “middle” of the support.

CONSTRUCTION 1.

$$\begin{aligned} \text{When } a_{K_n} - 1 > a_{K_1} + 1, & \begin{cases} f_{K+1}(a_{K_1}) = f_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_1} + 1) = f_K(a_{K_1} + 1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(x) = f_K(x) \quad \forall x \in \{a_{K_1} + 2, a_{K_1} + 3, \dots, a_{K_n} - 2\} \\ f_{K+1}(a_{K_n} - 1) = f_K(a_{K_n} - 1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_n}) = f_K(a_{K_n}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1} = 0 \text{ otherwise.} \end{cases} \\ \\ \text{When } a_{K_n} - 1 = a_{K_1} + 1, & \begin{cases} f_{K+1}(a_{K_1}) = f_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_1} + 1) = f_K(a_{K_1} + 1) + 2 \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1}(a_{K_n}) = f_K(a_{K_n}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ f_{K+1} = 0 \text{ otherwise.} \end{cases} \end{aligned}$$

Stop the sequence when \tilde{N}_T is such that $a_{T_n} - 1 < a_{T_1} + 1$.

Numerical Illustration 1: We illustrate Construction 1 with an example. Consider a random variable $\tilde{N}_K \in \{3, 4, 5, 7, 9\}$, with pmf $f_K(x) = \{0.1, 0.3, 0.2, 0.1, 0.3\}$ for $x = \{3, 4, 5, 7, 9\}$ respectively. To form \tilde{N}_{K+1} , Construction 1 requires $0.1 = \min\{0.1, 0.3\}$ of the probability mass at the

“ends” of the support to be re-allocated towards the “middle”, i.e., from “3” to “4”, and also from “9” to “8”. This results in $\tilde{N}_{K+1} \in \{4, 5, 7, 8, 9\}$, with pmf $f_{K+1}(x) = \{0.4, 0.2, 0.1, 0.1, 0.2\}$ for $x = \{4, 5, 7, 8, 9\}$. In the next step, applying the transformation from Construction 1 would lead to $\tilde{N}_{K+2} \in \{4, 5, 7, 8\}$, with pmf $f_{K+2}(x) = \{0.2, 0.4, 0.1, 0.3\}$ for $x = \{4, 5, 7, 8\}$.

With Construction 1, we create the sequence $\{\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_T\}$, beginning from an initial balking threshold distribution \tilde{N}_0 , with the corresponding cdf's $\{F_0, F_1, \dots, F_T\}$. Thus, we have a sequence of random variables that describe customer balking thresholds that are ordered in some sense. In the following Lemma 1, we show that T is finite and the distributions are mean-preserving spreads.

LEMMA 1. *Consider a sequence $\{\tilde{N}_K\}$ from Construction 1. (i) The sequence terminates at some finite $K = T$. (ii) $F_K \leq_{SMPS} F_{K-1}$ for $K \in \{1, 2, \dots, T\}$.*

Lemma 1 states that all the balking threshold distributions along the sequence built through Construction 1 have the same mean (i.e., they obey the mean preserving property). As long as the first balking threshold distribution is consistent, all the distributions in Construction 1 will be consistent. Furthermore, every succeeding distribution in the sequence is dominated (under the SMPS criterion) by the preceding distribution, i.e., every distribution in the sequence is followed by another distribution that has a lower “spread”.

We now show in Lemma 2, for all initial balking threshold distributions that have the same mean, the sequence *always* terminates at the same distribution \tilde{N}_T . Depending on the parameters of the initial distribution (support etc.), the number of steps taken to reach \tilde{N}_T may differ. Thus, T depends on the initial distribution, but \tilde{N}_T does not.

LEMMA 2. *Given any \tilde{N}_0 , the sequence $\{F_K\}$ terminates at the same F_T with the random variable $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}_0) \rfloor, \lceil \mathbb{E}(\tilde{N}_0) \rceil\}$ such that $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N}_0)$.*

Now that we have a sequence of random variables ordered SMPS by Construction 1, we can compare the performance metrics of the queueing system under different balking threshold distributions along the sequence. Using our notation introduced earlier, we denote $\lambda_{eff, \tilde{N}_K}, R_{\tilde{N}_K}, L_{\tilde{N}_K}$ and $W_{\tilde{N}_K}$ the effective arrival, the firm revenue, the average queue length, and the average waiting time, corresponding to \tilde{N}_K in the sequence of distributions $\{\tilde{N}_K\}_{K \geq 0}$.

LEMMA 3. *Let $\{\tilde{N}_K\}$ be any sequence from Construction 1. (i) $\lambda_{eff, \tilde{N}_K} < \lambda_{eff, \tilde{N}_{K+1}}$ ($R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$) for all ρ ; (ii) If $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right)$, then $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$; and (iii) If $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}-1}{a_{K_n}-2}\right)^2 + 4} - \frac{a_{K_n}-1}{a_{K_n}-2} \right)$, then $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$.*

Recall that Construction 1 builds a sequence of balking threshold distributions with decreasing spreads, while maintaining consistency (i.e., identical means). It follows from Lemma 3/(i) that, regardless of the traffic in the system, the revenues at the firm improve as the customers' balking thresholds become less spread-out (or narrower). This result is *not* due to Jensen's inequality.¹

Revenue improvements along the sequence emerge from the following two mechanisms: (i) The customer balking thresholds are gradually altered along the sequence in the construction which changes the long-run probabilities for *all* states. (ii) Along the construction path, the balking threshold increases for some customers, and decreases for some others. We prove that the throughput/revenue from increased joining of the customers with improved balking thresholds, compensates for the decreased joining of those customers with reduced balking thresholds. This is proven for *any* prior belief distribution.

Lemma 3/(ii) and (iii) provide *distribution-free* sufficient conditions for $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$, respectively. It is possible to derive stronger distribution-specific conditions for each inequality. Unlike the revenues, which always increase as the balking thresholds become less spread-out, the expected queue lengths and/or the expected waiting times can *increase* or *decrease*. We provide numerical examples below to support this observation.

Numerical Illustration 2: We explore the performance metrics as the balking threshold distribution \tilde{N}_K is transformed into \tilde{N}_{K+1} according to Construction 1. Since $\lambda_{eff, \tilde{N}_K} < \lambda_{eff, \tilde{N}_{K+1}}$ (from Lemma 3/(i)), following Little's Law, it is impossible to have $L_{\tilde{N}_K} > L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ at the same time. All three other cases are possible: (i) $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$, (ii) $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} > W_{\tilde{N}_{K+1}}$, and, (iii) $L_{\tilde{N}_K} > L_{\tilde{N}_{K+1}}$ and $W_{\tilde{N}_K} > W_{\tilde{N}_{K+1}}$.

For example, consider the random variable $\tilde{N}_K \in \{3, 4, 5\}$, such that its pmf $f_K(x) = \{0.2, 0.6, 0.2\}$ for $x = \{3, 4, 5\}$ respectively. Following Construction 1, we have \tilde{N}_{K+1} such that $Pr(\tilde{N}_{K+1} = 4) = 1$, which is also the last step. Let $\mu = 1$ in all cases below.

Case (i): When $\rho = 0.4$, $L_{\tilde{N}_K} = 0.609 < L_{\tilde{N}_{K+1}} = 0.615$ and $W_{\tilde{N}_K} = 1.551 < W_{\tilde{N}_{K+1}} = 1.562$. ($\lambda_{eff, \tilde{N}_K} = 0.392 < \lambda_{eff, \tilde{N}_{K+1}} = 0.394$.)

Case (ii): When $\rho = 0.825$, $L_{\tilde{N}_K} = 1.617 < L_{\tilde{N}_{K+1}} = 1.621$ and $W_{\tilde{N}_K} = 2.265 > W_{\tilde{N}_{K+1}} = 2.262$. ($\lambda_{eff, \tilde{N}_K} = 0.714 < \lambda_{eff, \tilde{N}_{K+1}} = 0.717$.)

¹ Consider a distribution \tilde{N}_0 with cdf F_0 , pmf f_0 and some integer mean $\mathbb{E}(\tilde{N}_0)$. Jensen's inequality would imply $R_{\mathbb{E}(\tilde{N}_0)} > \sum_{n=0}^{\infty} f_0(n)R_n$ where $R_{\mathbb{E}(\tilde{N}_0)}$ and R_n are revenues when *all* customers use the balking threshold $\mathbb{E}(\tilde{N}_0)$ and n respectively. Lemma 3 states that $R_{\mathbb{E}(\tilde{N}_0)} > R_{\tilde{N}_0} = p\mu(1 - \pi_0)$ where $\pi_0 = 1 / \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_0(n)\right)$ (from Equation (1)) which is not related to $\sum_{n=0}^{\infty} f_0(n)R_n$.

Case (iii): When $\rho = 0.9$, $L_{\tilde{N}_K} = 1.793 > L_{\tilde{N}_{K+1}} = 1.790$ and $W_{\tilde{N}_K} = 2.380 > W_{\tilde{N}_{K+1}} = 2.368$. ($\lambda_{eff, \tilde{N}_K} = 0.753 < \lambda_{eff, \tilde{N}_{K+1}} = 0.756$.)

When revenue/throughput is higher under some change in balking threshold distribution, one would expect the congestion to also be higher because more customers join. However, we find scenarios in which revenue/throughput increases while congestion decreases. According to Lemma 3/(ii) and (iii), this occurs only when traffic is sufficiently large (i.e., large enough ρ). Theorem 2 formalizes this observation.

Having illustrated the comparative statics for the sequence of distributions in Lemma 3, we can now compare the performance metrics of (any) initial distribution with the terminal distribution. This is captured in Theorem 2. It turns out that when customers' balking thresholds become less spread-out, the firm always improves its revenues. On the other hand, customers have to wait longer on average, if the traffic is smaller than some threshold level.

THEOREM 2. *Let \tilde{N} be any balking threshold distribution and \tilde{N}_T be the last term from Construction 1 initiated at $\tilde{N}_0 = \tilde{N}$. Then, (i) $\lambda_{eff, \tilde{N}} < \lambda_{eff, \tilde{N}_T}$ ($R_{\tilde{N}} < R_{\tilde{N}_T}$) for all ρ ; (ii) $\exists \rho_L$ s.t. $L_{\tilde{N}} < L_{\tilde{N}_T} \forall \rho \leq \rho_L$; and (iii) $\exists \rho_W$ s.t. $W_{\tilde{N}} < W_{\tilde{N}_T} \forall \rho \leq \rho_W$.*

Theorem 2 indicates that revenue at the server always improves when balking thresholds become less spread-out. Essentially, some customers with high balking threshold become less patient, and some others with low balking thresholds become more patient. Due to PASTA property, the system occupancy when customers arrive is more likely to be low (recall $\rho < 1$). As a result, there is increased joining of customers leading to higher revenues.

The effect on congestion and waiting times depends on the traffic ρ . When the traffic intensity ρ is small, balking thresholds that are *less* spread-out can increase congestion (L or W). This could be understood through externalities imposed by joining/balking customers.

At low queue-lengths, more customers join (as balking thresholds become less spread-out) causing increased negative externalities for future arrivals. At high queue-lengths, less customers join decreasing the negative externalities at those states. The net effect of the negative externalities imposed depends on the likelihood of the low queues lengths to high queue lengths.

When ρ is small, the queue mostly resides at low states and visits higher states less often. Thus when ρ is small, the increased negative externalities imposed by customers joining at low states exceed any benefit from reduced negative externalities at high states. As a result, the expected wait times and queue lengths are higher when the beliefs are less spread-out.

When ρ is large, the higher states are relatively more likely to be visited (as opposed to when ρ is small). Thus, it is possible that the benefits accrued at higher states can overcome the negative externalities imposed at lower states. For instance, see Cases (ii) and (iii) of Numerical Illustration 2. Thus, in low traffic, as balking thresholds become more accurate, customers wait longer and suffer higher disutility. This result is intriguing because serving the same population, a faster server announcing its true service rate is more likely to result in welfare loss due to increased congestion.

Lower Bounds: We can use the analytical properties of the bounds along the sequence to derive *distribution-free* bounds (with respect to \tilde{N}) on ρ_L and ρ_W that hold for any arbitrary customer balking threshold distribution. The lower bounds $\underline{\rho}_L$ and $\underline{\rho}_W$ are such that $\rho_L \geq \underline{\rho}_L = 0.5$ and $\rho_W \geq \underline{\rho}_W = 0.414$ respectively. We defer the details of the derivation to the appendix.

4. Impact of Revealing Service Information

We now calibrate the impact of a firm revealing its true service rate μ , by comparing revenues and system congestion under two balking threshold distributions corresponding to the case when the firm reveals or does not reveal its service information.

When customers are uninformed, they may have arbitrary beliefs $\tilde{\mu}$ over the service rate, and the balking thresholds are distributed according to $\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor$. When the firm reveals its true service rate, all customers will follow the same true balking threshold $N = \lfloor \frac{\mu(v-p)}{c} \rfloor$. Since the threshold distributions (that emerge from arbitrary beliefs) could be arbitrary, it is unclear when a firm should reveal its service rate. We examine this question in the following proposition.

PROPOSITION 1. *When $\mathbb{E}(\tilde{N}) \leq N$, the firm's revenue increases on revealing service information (i.e., $R \uparrow$). In addition, when traffic ρ is small, the average queue length and the average customer waiting time in the system both increase on revelation ($L, W \uparrow$).*

From the Proposition, we find that when customer balking distributions are pessimistic or even consistent, i.e., $\mathbb{E}(\tilde{N}) \leq N$, it is always in the firm's interest to reveal its service rate as its throughput and revenues improve. Yet, when ρ is small, the system congestion (the average queue length and the average customer waiting time) also increases on customers knowing the true information.

To intuit the result, consider a population with balking threshold distribution \tilde{N} . Let $\{\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_T\}$ be a sequence of balking threshold distributions from Construction 1 starting with $\tilde{N}_0 = \tilde{N}$. Theorem 2 states that $R \uparrow$ and $L, W \uparrow$ (for small ρ) when customers adopt \tilde{N}_T instead of \tilde{N} . Now suppose the balking thresholds are pessimistic or consistent, i.e., $\mathbb{E}(\tilde{N}) \leq N$, it then follows that $\tilde{N}_T \leq_{st} N$. So by Theorem 1, upon revealing, we have $R \uparrow$ and $L, W \uparrow$. Thus, combining Theorems 1 and 2, when $\mathbb{E}(\tilde{N}) \leq N$, we have $R \uparrow$ and $L, W \uparrow$ (for small ρ).

Now suppose that customer balking thresholds have optimism bias ($N < \mathbb{E}(\tilde{N})$). Again, we construct the sequence $\{\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_T\}$ using Construction 1 starting with $\tilde{N}_0 = \tilde{N}$. We have $R_{\tilde{N}} < R_{\tilde{N}_1} < \dots < R_{\tilde{N}_T}$ by Theorem 2. On the other hand, $N < \mathbb{E}(\tilde{N})$ implies $N <_{st} \tilde{N}_T$, so by Theorem 1 we have $R_N < R_{\tilde{N}_T}$. Depending on \tilde{N} , we may have $R_N < R_{\tilde{N}}$ or $R_N > R_{\tilde{N}}$. Recall that $\{\tilde{N}_0, \tilde{N}_1, \dots, \tilde{N}_T\}$ is a sequence of distributions that have progressively lower spreads. We can conclude that when the customer balking thresholds are optimistic, the firm may still reveal its service information as long as it observes high variance in its distribution. We provide numerical examples to support this observation in §4.2.

Remark: The results in Proposition 1 are distribution-free and also parameter-free, i.e., it is sufficient for the firm to only know that the balking threshold distributions are pessimistic or consistent, before the decision to reveal true information is made. The exact distribution itself does not influence the decision to reveal information.

4.1. Service Rate Beliefs vs. Balking Thresholds

At this juncture, the reader may note that all our preceding analysis is based on balking thresholds. In Proposition 1, we have customer pessimism in terms of the balking thresholds rather than the internally held beliefs on service-rate. Observe that balking thresholds could be estimated econometrically by the service provider, by observing the long-run time averages of balking at different states. However, the distribution of *exact* service rate beliefs cannot be elicited by observing the joining/balking actions, since (infinitely) many service rate beliefs could result in an identical balking threshold. Suppose the service rate beliefs could be identified. Lemma 4 below characterizes the correspondence between bias (pessimism/optimism) in service-rate beliefs, to the bias in the resulting balking threshold distribution.

LEMMA 4. *For any given service rate belief distribution $\tilde{\mu}$ and some true service rate μ , we have (i) $\mu \geq \mathbb{E}(\tilde{\mu}) + \frac{c}{v-p} \Rightarrow \mathbb{E}(\tilde{N}) < N$; (ii) $\mu \leq \mathbb{E}(\tilde{\mu}) - \frac{c}{v-p} \Rightarrow \mathbb{E}(\tilde{N}) > N$; and (iii) $\forall x \in (0, 1]$, $|\mathbb{E}(\tilde{\mu}) - \mu| < x \cdot \frac{c}{v-p} \Rightarrow |\mathbb{E}(\tilde{N}) - N| < 1 + x$, especially when $\mathbb{E}(\tilde{\mu}) = \mu$, $|\mathbb{E}(\tilde{N}) - N| < 1$.*

Lemma 4/(i) indicates that if consumer service belief distribution is *sufficiently pessimistic* (i.e., $\mathbb{E}(\tilde{\mu}) \leq \mu - \frac{c}{v-p}$), then the corresponding balking threshold distribution will also be pessimistic. (Proposition 1 applies.) Similarly, from 4/(ii), if consumer belief distribution is *sufficiently optimistic* (i.e., $\mathbb{E}(\tilde{\mu}) \geq \mu + \frac{c}{v-p}$), then the balking threshold distribution will also be optimistic. If the bias in service rate beliefs is sufficiently small (i.e., $|\mathbb{E}(\tilde{\mu}) - \mu| < \frac{c}{v-p}$), the consumer balking threshold distribution may be either slightly pessimistic, slightly optimistic, or consistent, but its bias $|\mathbb{E}(\tilde{N}) - N|$ is strictly bounded as provided in Lemma 4/(iii). In all cases, the difference between

bias in service rate beliefs and bias in balking thresholds decreases in value of service relative to the waiting cost (i.e., $\frac{c}{v-p} \downarrow$). When $\frac{c}{v-p} \rightarrow 0$, the difference in biases disappears.

Observing Lemma 4/(iii), it is possible that the “internally-held” service-rate belief distribution is slightly pessimistic (i.e. $\mu - \frac{c}{v-p} < \mathbb{E}(\tilde{\mu}) < \mu$), whereas the resulting balking threshold distribution is optimistic and vice-versa. We provide numerical examples below to further illustrate the divergence.

Numerical Illustration 3: In both examples below, we assume that the customers are unaware of the provider’s true service rate, μ , and their service rate beliefs are uniformly distributed over $[2, 8]$ per unit time (the time units for μ or $\tilde{\mu}$ omitted henceforth), i.e., $\tilde{\mu} \sim U[2, 8]$ with mean $\mathbb{E}(\tilde{\mu}) = 5$.

Example (i): We set the service value $v = \$8$, price $p = \$2$, and customer linear waiting cost at $c = \$4/\text{time unit}$. As a result, customers’ balking threshold, $\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor = \lfloor 3\tilde{\mu}/2 \rfloor$, is a discrete uniform distribution taking values $\{3, 4, \dots, 11\}$ with mean $\mathbb{E}(\tilde{N}) = 7$.

If service rate belief distribution is ...	$\mathbb{E}(\tilde{\mu}) = 5$	$\mathbb{E}(\tilde{N}) = 7$	Balking threshold distribution is ...
sufficiently pessimistic defined by Lemma 4/(i)	$\mu \geq 5\frac{2}{3}$	$\Rightarrow N \geq 8$	pessimistic
slightly pessimistic, slightly optimistic or consistent defined by Lemma 4/(iii)	$5\frac{2}{3} > \mu \geq 5\frac{1}{3}$	$\Rightarrow N = 8$	
	$5\frac{1}{3} > \mu \geq 4\frac{2}{3}$	$\Rightarrow N = 7$	consistent
	$4\frac{2}{3} > \mu > 4\frac{1}{3}$	$\Rightarrow N = 6$	optimistic
sufficiently optimistic defined by Lemma 4/(ii)	$\mu \leq 4\frac{1}{3}$	$\Rightarrow N \leq 6$	

Table 1 Bias in Service Rate Beliefs and Balking Thresholds when $\tilde{\mu} \sim U[2, 8]$ and $\frac{c}{v-p} = \frac{2}{3}$.

In Table 1, we illustrate the correspondence between the service rate belief distribution and the balking threshold distribution. We cover the three parts described in Lemma 4 by looking at various possible value of the “true” service rate μ . Specifically, as stated before, if the service rate belief distribution is slightly pessimistic, slightly optimistic or consistent, the corresponding balking threshold distribution can be consistent, (slightly) pessimistic or (slightly) optimistic.

Example (ii): For the second example, we set the service value $v = \$8$, price $p = \$2$, and customer linear waiting cost at $c = \$8/\text{time unit}$. Customers’ balking threshold distribution, $\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor = \lfloor 3\tilde{\mu}/4 \rfloor$, thus takes values $\{1, 2, 3, 4, 5\}$ with mean $\mathbb{E}(\tilde{N}) = 3\frac{2}{9}$.

If service rate belief distribution is ...	$\mathbb{E}(\tilde{\mu}) = 5$	$\mathbb{E}(\tilde{N}) = 3\frac{2}{9}$	Balking threshold distribution is ...
sufficiently pessimistic defined by Lemma 4/(i)	$\mu \geq 6\frac{1}{3}$	$\Rightarrow N \geq 4$	pessimistic
slightly pessimistic, slightly optimistic or consistent defined by Lemma 4/(iii)	$6\frac{1}{3} > \mu \geq 5\frac{1}{3}$	$\Rightarrow N = 4$	
	$5\frac{1}{3} > \mu \geq 4$	$\Rightarrow N = 3$	optimistic
	$4 > \mu > 3\frac{2}{3}$	$\Rightarrow N = 2$	
sufficiently optimistic defined by Lemma 4/(ii)	$\mu \leq 3\frac{2}{3}$	$\Rightarrow N \leq 2$	

Table 2 Bias in Service Rate Beliefs and Balking Thresholds when $\tilde{\mu} \sim U[2, 8]$ and $\frac{c}{v-p} = \frac{4}{3}$.

As the value of $\frac{c}{v-p}$ doubles in this example, difference in biases in service rate beliefs and balking threshold distribution increases. For instance, when $\mu = 5.25$ in Table 2, note that the service rate beliefs are pessimistic and the resulting threshold distribution is optimistic.

Remark: Note that the distributional differences between $\tilde{\mu}$ and \tilde{N} , are not due to assumptions in our model, but a direct outcome of discrete join/balk decisions in observable queues. Thus, revealing the true service rate information, has the same impact as revealing the corresponding “true” balking threshold information. In other words, differences in biases do not impact our findings in Proposition 1, as long as the balking threshold distribution is characterized accurately.

4.2. Revenue and Welfare Effects of True Information

We now study the impact of revelation of the service information (or the true balking threshold) on revenue and welfare by exploring specific cases of $M/M/1$ queues under different biases. In all cases, we use the same parameters v, p, c and service rate belief distribution $\tilde{\mu}$ as in Example (i) of Numerical Illustration 3 in §4.1, i.e., $\tilde{\mu} \sim U[2, 8]$ and $\frac{c}{v-p} = \frac{2}{3}$.

We examine three scenarios highlighted in the right-most column of Table 1 (Balking threshold distribution is ...) where the resulting threshold N from announcing the true service rate is (i) greater than, (ii) equal to, or (iii) less than $\mathbb{E}(\tilde{N})$. These three instances correspond to (i) $\mu \geq 5\frac{1}{3}$, sufficient or slight pessimism, (ii) $\mu \in [4\frac{2}{3}, 5\frac{1}{3})$, slight optimism/slight pessimism/consistent, (iii) $\mu < 4\frac{2}{3}$, sufficient or slight optimism in customer service rate beliefs. For each case, we examine the firm’s revenue, the average queue length and the average customer waiting time, as well as consumer welfare and social welfare.

The first line in each table that follows is corresponding to some initial situation when customers are unaware of the firm’s service rate information (and adopt balking threshold distribution \tilde{N}); the last line of each table corresponds to the situation when the firm reveals the service rate information to its customers (customers thus adopt balking threshold distribution N). All rows in between the first and the last rows communicate the terms in the sequence in Construction 1. The percentage change in a parameter (compared to the original beliefs \tilde{N} , first line) is noted in parenthesis.

Pessimistic Balking Thresholds: Assume that the true service rate is $\mu = 6$ and the arrival rate is $\lambda = 5$. Note that $N = 9 > \mathbb{E}(\tilde{N}) = 7$. Since $\mu \geq \mathbb{E}(\tilde{\mu}) + \frac{c}{v-p} = 5\frac{2}{3}$, the customers have sufficiently pessimistic service rate beliefs.

Balking Thresholds	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	8.75	1.86	0.42	18.81	27.56
Construction: \tilde{N}_1	8.87 (+1.41%)	1.94 (+4.32%)	0.44 (+2.87%)	18.86 (+0.26%)	27.73 (+0.62%)
\tilde{N}_2	9.03 (+3.23%)	2.08 (+11.94%)	0.46 (+8.44%)	18.77 (-0.21%)	27.80 (+0.88%)
\tilde{N}_3	9.17 (+4.88%)	2.24 (+20.47%)	0.49 (+14.86%)	18.57 (-1.28%)	27.75 (+0.68%)
$\tilde{N}_4 = \tilde{N}_T$	9.23 (+5.47%)	2.29 (+23.35%)	0.50 (+16.95%)	18.52 (-1.58%)	27.74 (+0.66%)
Informed N	9.52 (+8.83%)	2.84 (+52.78%)	0.60 (+40.38%)	17.21 (-8.52%)	26.73 (-3.01%)

In this case, revealing the true service rate increases the firm's revenue by 8.83% but also increases the average queue length by 52.78% and the average waiting time by 40.38%. The firm thus benefits from revealing its service rate information (in line with Proposition 1), but the increased benefit is not sufficient to overcome the loss in consumer welfare (-8.52%). As a result, the overall social welfare drops by 3.01%.

Consistent Balking Thresholds: Now assume that $\mu \in [4\frac{2}{3}, 5\frac{1}{3})$ and $\lambda = 4$. Therefore, customers have slightly pessimistic, slightly optimistic or possibly consistent service rate beliefs (i.e., $|\mathbb{E}(\tilde{\mu}) - \mu| < \frac{c}{v-p}$). However, $N = \mathbb{E}(\tilde{N}) = 7$. Thus the resulting balking threshold distribution is consistent.

Balking Thresholds	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	7.08 (+0.00%)	1.74 (+0.00%)	0.49 (+0.00%)	14.29 (+0.00%)	21.37 (+0.00%)
Construction \tilde{N}_1	7.18 (+1.43%)	1.82 (+4.37%)	0.51 (+2.89%)	14.29 (+0.00%)	21.47 (+0.47%)
\tilde{N}_2	7.31 (+3.24%)	1.95 (+11.89%)	0.53 (+8.37%)	14.15 (-0.97%)	21.46 (+0.43%)
\tilde{N}_3	7.43 (+4.86%)	2.09 (+20.19%)	0.56 (+14.62%)	13.92 (-2.61%)	21.34 (-0.13%)
$\tilde{N}_4 = \tilde{N}_T$	7.47 (+5.46%)	2.14 (+23.12%)	0.57 (+16.75%)	13.84 (-3.15%)	21.31 (-0.30%)
Informed $N = \tilde{N}_T$	7.47 (+5.46%)	2.14 (+23.12%)	0.57 (+16.75%)	13.84 (-3.15%)	21.31 (-0.30%)

In this example, revealing the true service rate still improves revenues (by 5.46%) in line with Proposition 1. However, the average queue length and the average waiting time both increase significantly (by 23.12% and 16.75% respectively). The firm benefits from revealing the service rate, almost fully at the expense of consumer welfare (-3.15%), but the overall social welfare does not fall significantly (-0.30%) due to the increase in throughput (i.e., number of customers served).

Optimistic Balking Thresholds: Finally, let $\mu = 4$ and $\lambda = 3$. Note that $N = 6 < \mathbb{E}(\tilde{N}) = 7$ and $\mu \leq \mathbb{E}(\tilde{\mu}) - \frac{c}{v-p} = 4\frac{1}{3}$ (and hence, the population has sufficiently optimistic service rate beliefs).

Balking Thresholds	Firm Revenue	Avg. Queue Length	Avg. Waiting Time	Consumer Welfare	Social Welfare
Uninformed $\tilde{N} = \tilde{N}_0$	5.40 (+0.00%)	1.57 (+0.00%)	0.58 (+0.00%)	9.93 (+0.00%)	15.33 (+0.00%)
Construction: \tilde{N}_1	5.48 (+1.45%)	1.64 (+4.39%)	0.60 (+2.90%)	9.89 (-0.41%)	15.37 (+0.25%)
\tilde{N}_2	5.58 (+3.22%)	1.75 (+11.65%)	0.63 (+8.17%)	9.72 (-2.11%)	15.30 (-0.23%)
\tilde{N}_3	5.66 (+4.76%)	1.88 (+19.48%)	0.66 (+14.04%)	9.48 (-4.54%)	15.14 (-1.26%)
$\tilde{N}_4 = \tilde{N}_T$	5.69 (+5.34%)	1.92 (+22.37%)	0.68 (+16.16%)	9.39 (-5.43%)	15.08 (-1.64%)
Informed N	5.57 (+3.03%)	1.70 (+8.31%)	0.61 (+5.13%)	9.90 (-0.31%)	15.46 (-0.86%)

In this case, even though the customers are (incorrectly) optimistic about the service rate, revealing the true service rate may still increase the firm's revenue (by 3.03%). Examining the second column of the table (firm's revenue column) reveals that when customers' optimistic balking thresholds are more dispersed, it is beneficial for the firm to reveal service rate.

Observing further, we see that if customers' balking threshold distribution is characterized by \tilde{N} or \tilde{N}_1 , the firm increases its revenue from revealing the true service rate (whereas customers suffer longer queues under full information). On the other hand, if customer thresholds are less dispersed,

(for e.g., \tilde{N}_3), the firm does not gain from revealing its service rate. However, customers would be better off with the revealed information (through lower expected queue lengths and wait times).

In summary:

- Revenues generally improve with more information but welfare effects are mixed. The gains in revenues are usually, but not always, lower than consumer welfare loss. Thus, social welfare can reduce, as a consequence of more information in the system.
- When is it possible that both the firm revenues and consumer welfare improve upon service information revelation? Such welfare improvements only occur when the traffic ρ is high and customers' initial beliefs are almost deterministic. One such example is given by Case (iii) of Numerical Illustration 2 in §3.2 where $\lambda_{eff} \uparrow$ ($R \uparrow$) and $L, W \downarrow$.

5. Conclusions and Future Research.

Customers often join queues with very limited information. Most of the literature has assumed that service parameters (specifically, service rate μ) that influence the joining behavior as common knowledge. However, customers cannot always fully characterize these service parameters; sometimes even the calculation of mean service time may require repeated sampling or collection of data. In such queues without information, not much is known on how revenues and welfare are impacted, when a firm reveals its service information (specifically, service rate). Our paper seeks to fill this gap.

Our approach to solving the information problem is distribution-free. Using our general but intuitive approach, we calibrate the impact of information revelation on the performance of the queueing system, without any restrictions on the distribution of the initial customer beliefs. We can apply the results from our general model on belief structures that may emerge from specific behaviors such as Quantal-response based bounded rationality (Luce (1959), Su (2008), Huang et al. (2013)), out-of-queue learning through sampling from past experiences (Xu et al. 2007), anecdotal reasoning (Huang and Chen 2014) and other cognitive biases to characterize their effects on revenues and consumer welfare.

Insights and Implications: We find that the impact of service rate information on congestion and welfare is mixed. Even though a firm's revenue improves on announcing its service rate under certain conditions (e.g., when the balking threshold distribution is pessimistic or consistent), the congestion levels such as the average queue length or average wait time increase. As a result, consumer welfare worsens with more information, despite the increased market coverage. In fact,

consumer welfare loss can exceed revenue improvements at the firm. Thus, social welfare can fall with provision of correct public information.

Hence, intriguingly, with informational uncertainty, social welfare typically improves in queues, compared to when customers have full information. When left to their own devices, with full information, more customers join a queue than what is socially optimal. In Naor (1969), tolls/taxes are levied to control the joining population to improve welfare. Likewise, we find that lack of information acts as an *information tax* that deters admission, which could lead to improved welfare.

The impact of additional information changes with the service capacity, given a market size. Consumer welfare likely worsens in the case when a fast server reveals its service rate, compared to the case when a slower server reveals its rate. Thus, customers are more likely to be worse off if a *faster* firm with higher service capacity reveals its service information.

Our findings have several implications for queue management policies in practice. In primary healthcare settings where the access to service providers is important, revealing the capacity information can lead to an increased customer access to the queue (i.e., more customers will visit the service provider). Nevertheless, customers will observe longer queues on average, and also suffer a higher dis-utility in waiting time on average. Thus when there is a significant impetus on treating admitted patients quickly, as in some emergency room settings, revealing the service information may lead to increased crowding and worsen the average wait times. Furthermore, this effect is exacerbated for a facility that has ample service capacity. So, the decision to reveal the service information in aforementioned settings depends critically on the trade-offs between improved access and increased congestion.

Further Considerations: Our model has some limitations and is amenable to further extensions. We considered systems with $\lambda < \mu$: The effect of information revelation to misinformed populations in overloaded service systems is still open. We believe that much of our results of the first-order stochastics continue to hold, but it appears that sequencing beliefs under higher stochastic orders is more complex. Extensions to multiple server settings preserve our theoretical findings, and proofs are available from the authors.

Price is exogenous in our model. An interesting key question would be to examine how a firm would make joint decisions over price and service information revelation.

Research directions related to information and learning in queues appear promising. For instance, it is unclear how consumers process limited information when the queue lengths are not visible, or when others' waiting costs may vary. In unobservable queue settings, every customer does not only needs to use his own belief over the service parameters, but also needs to develop beliefs

on other customers' actions or beliefs, based on limited information. In such cases, the timing of delay announcements (such as Allon et al. (2014)), or shading the information to be provided play a critical role in customer decisions. Finally, our model has no reneging. In real settings, customers may often accrue information and correct their beliefs as they wait in queues, leading to abandonment behavior and future retrials. Such within-queue learning and rational abandonment behaviors are fruitful directions of future research endeavors.

Appendix

Proof of Lemma 1:

(i) By construction, the entire probability mass at one end of the distribution is transferred towards the middle of the support. As a result, the range of the random variable \tilde{N}_{K+1} is a strict subset of the range of \tilde{N}_K . Specifically, $a_{(K+1)_1} = a_{K_1} + 1$ if $f_K(a_{K_1}) \leq f_K(a_{K_n})$ and $a_{(K+1)_N} = a_{K_N} - 1$ if $f_K(a_{K_1}) \geq f_K(a_{K_n})$. The length of the range of \tilde{N}_K , $|a_{K_n} - a_{K_1}|$, is strictly decreasing in K . Within a finite number of steps, for some time $K = T$, the length will be less than 2. When $a_{T_n} - 1 < a_{T_1} + 1$, the process stops. Thus, T is finite.

(ii) We show that $F_{K+1} \leq_{SMPs} F_K$. Let $a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4}$ in Definition 2 be $a_{K_1}, a_{K_1} + 1, a_{K_n} - 1$ and a_{K_n} respectively. $f_{K+1} = f_K$ for all but these four points. Define $\gamma_{i_k} = f_K(a_{i_k}) - f_{K+1}(a_{i_k})$ for $k = 1, 2, 3, 4$. Then, $\gamma_{i_1} = -\gamma_{i_2} = -\gamma_{i_3} = \gamma_{i_4} = \min\{f_K(a_{K_1}), f_K(a_{K_n})\} > 0$. Moreover, $\sum_{k=i}^4 a_{i_k} \gamma_{i_k} = [a_{K_1} - (a_{K_1} + 1) - (a_{K_n} - 1) + a_{K_n}] \min\{f_K(a_{K_1}), f_K(a_{K_n})\} = 0 \cdot \min\{f_K(a_{K_1}), f_K(a_{K_n})\} = 0$. \square

Proof of Lemma 2:

Suppose that \tilde{N}_T has two elements which are not consecutive. Then, it must be that $a_{T_1} + 1 < a_{T_n}$. By Construction 1, then sequence is not completed, which contradicts the definition of T . Else, suppose that \tilde{N}_T has three or more elements. Again, it must be that $a_{T_1} + 1 < a_{T_n}$, and hence, the sequence in Construction 1 is incomplete, which contradicts the definition of T . Therefore, \tilde{N}_T can either take a single value or two consecutive values. Case (i): When if $\mathbb{E}(\tilde{N}_0)$ is an integer, since the transformation is mean preserving, we have \tilde{N}_T is a singleton with $\tilde{N}_T = \mathbb{E}(\tilde{N}_0) = \lfloor \mathbb{E}(\tilde{N}_0) \rfloor = \lceil \mathbb{E}(\tilde{N}_0) \rceil$. Case (ii): When $\mathbb{E}(\tilde{N}_0)$ is not an integer, \tilde{N}_T cannot be a singleton. Thus, \tilde{N}_T takes on two consecutive values. Since the transformation in Construction 1 is mean-preserving with $\mathbb{E}(\tilde{N}_0)$, we have $\mathbb{E}(\tilde{N}_T) = \mathbb{E}(\tilde{N}_0)$. Then we must have $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}_0) \rfloor, \lceil \mathbb{E}(\tilde{N}_0) \rceil\}$, with $\Pr(\tilde{N}_T = \lfloor \mathbb{E}(\tilde{N}_0) \rfloor) \lfloor \mathbb{E}(\tilde{N}_0) \rfloor + \Pr(\tilde{N}_T = \lceil \mathbb{E}(\tilde{N}_0) \rceil) \lceil \mathbb{E}(\tilde{N}_0) \rceil = \mathbb{E}(\tilde{N}_0)$. It is also clear that the distribution of \tilde{N}_T is independent of the distribution of \tilde{N}_0 . \square

Proof of Lemma 3:

(i) For $j \in \{K, K+1\}$, recall from (4) that $R_{\tilde{N}_j} = p\lambda_{eff, \tilde{N}_j} = p\mu(1 - \pi_0)$ where $\pi_0 = 1 / \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_j(n)\right)$. It is thus sufficient to show that

$$\sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i=1}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n). \quad (7)$$

To verify (7), our strategy is to form a partition of $i \in \{0, 1, 2, \dots\}$ based on the sign of $\rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) - \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$. We specifically focus on terms that make the product $\prod_{n=0}^{i-1} \bar{F}_K(n)$, namely $\bar{F}_K(n)$. Since $\bar{F}_K(n) = f_K(n+1) + f_K(n+2) + \dots$, applying Construction 1, we have

$$\begin{aligned}
\bar{F}_{K+1}(0) &= \bar{F}_K(0) = 1, \\
\bar{F}_{K+1}(1) &= \bar{F}_K(1) = 1, \\
&\vdots \\
\bar{F}_{K+1}(a_{K_1} - 1) &= \bar{F}_K(a_{K_1} - 1) = 1, \\
\bar{F}_{K+1}(a_{K_1}) &= \bar{F}_K(a_{K_1}) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \in (0, 1], \\
\bar{F}_{K+1}(a_{K_1} + 1) &= \bar{F}_K(a_{K_1} + 1) \in (0, 1), \\
\bar{F}_{K+1}(a_{K_1} + 2) &= \bar{F}_K(a_{K_1} + 2) \in (0, 1), \\
&\vdots \\
\bar{F}_{K+1}(a_{K_n} - 2) &= \bar{F}_K(a_{K_n} - 2) \in (0, 1), \\
\bar{F}_{K+1}(a_{K_n}) &= \bar{F}_K(a_{K_n}) = 0, \\
\bar{F}_{K+1}(a_{K_n} + 1) &= \bar{F}_K(a_{K_n} + 1) = 0, \\
&\vdots
\end{aligned} \tag{8}$$

Thus, using transformation of \tilde{N}_K to \tilde{N}_{K+1} in Construction 1, we see that $\bar{F}_K(n)$ differs from $\bar{F}_{K+1}(n)$ at only two points, specifically $n = a_{K_1}$ and $n = a_{K_n} - 1$. In order to show (7), we verify, as an intermediate step, that $\bar{F}_{K+1}(a_{K_1}) \cdot \bar{F}_{K+1}(a_{K_n} - 1) < \bar{F}_K(a_{K_1}) \cdot \bar{F}_K(a_{K_n} - 1)$. We have

$$\begin{aligned}
&\bar{F}_{K+1}(a_{K_1}) \cdot \bar{F}_{K+1}(a_{K_n} - 1) = [\bar{F}_K(a_{K_1}) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\}][\bar{F}_K(a_{K_n} - 1) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\}] \\
&= \bar{F}_K(a_{K_1})\bar{F}_K(a_{K_n} - 1) + \min\{f_K(a_{K_1}), f_K(a_{K_n})\}[\bar{F}_K(a_{K_n} - 1) - \bar{F}_K(a_{K_1}) - \min\{f_K(a_{K_1}), f_K(a_{K_n})\}] \\
&< \bar{F}_K(a_{K_1})\bar{F}_K(a_{K_n} - 1) \text{ since } \bar{F}_K(a_{K_n} - 1) \leq \bar{F}_K(a_{K_1}).
\end{aligned} \tag{9}$$

Now we define $\mathcal{S}_1 \triangleq \{1, 2, \dots, a_{K_1}\}$; $\mathcal{S}_2 \triangleq \{a_{K_1} + 1, a_{K_1} + 2, \dots, a_{K_n} - 1\}$; $\mathcal{S}_3 \triangleq \{a_{K_n}\}$; and $\mathcal{S}_4 \triangleq \{a_{K_n} + 1, a_{K_n} + 2, \dots\}$. $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and \mathcal{S}_4 then form a partition of the space $\{1, 2, 3, \dots\}$. Our goal (7) is equivalent to

$$\sum_{i \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n).$$

From (8) and (9) we have $\forall i \in \mathcal{S}_1 : \prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) = 1$; $\forall i \in \mathcal{S}_2 : \prod_{n=0}^{i-1} \bar{F}_K(n) < \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$; $\forall i \in \mathcal{S}_3 : \prod_{n=0}^{i-1} \bar{F}_K(n) > \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$; and $\forall i \in \mathcal{S}_4 : \prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) = 0$.

It is clear that \mathcal{S}_1 and \mathcal{S}_4 are collection of the indices i where $\prod_{n=0}^{i-1} \bar{F}_K(n) = \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$. Hence, to prove (7), it suffices to show that $\sum_{i \in \mathcal{S}_2 \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_2 \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$. As $\rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)$ for all $i \in \mathcal{S}_2$, the inequality will hold, if there exists some $\mathcal{S}_{2'} \subseteq \mathcal{S}_2$ such that

$$\sum_{i \in \mathcal{S}_{2'} \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \mathcal{S}_{2'} \cup \mathcal{S}_3} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n). \tag{10}$$

On the other hand, the existence of \tilde{N}_{K+1} guarantees that $a_{K_n} - 1 \geq a_{K_1} + 1$ and $a_{K_1} \geq 1$ so there exists

at least one index in \mathcal{S}_2 (i.e., $i = a_{K_1} + 1$). There is only one element in \mathcal{S}_3 (i.e., $i = a_{K_n}$). Let us define $\mathcal{S}_{2'} \triangleq \{a_{K_1} + 1\}$. So $\mathcal{S}_{2'} \cup \mathcal{S}_3 = \{a_{K_1} + 1, a_{K_n}\}$. Inequality (10) is therefore equivalent to

$$\sum_{i \in \{a_{K_1}+1, a_{K_n}\}} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) < \sum_{i \in \{a_{K_1}+1, a_{K_n}\}} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n),$$

$$\Leftrightarrow \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}-1} \bar{F}_K(n) + \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) < \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}-1} \bar{F}_{K+1}(n) + \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n).$$

And the last condition is true because

$$\begin{aligned} & \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) - \rho^{a_{K_n}} \prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n) < \rho^{a_{K_n}} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \\ & \leq \rho^{a_{K_n}} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} < \rho^{a_{K_1}+1} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \\ & = \rho^{a_{K_1}+1} \min\{f_K(a_{K_1}), f_K(a_{K_n})\} \prod_{n=0}^{a_{K_1}-1} \bar{F}_K(n) = \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}-1} \bar{F}_{K+1}(n) - \rho^{a_{K_1}+1} \prod_{n=0}^{a_{K_1}-1} \bar{F}_K(n). \end{aligned}$$

Therefore, inequality (10)-(7) all hold by backward induction, and $\lambda_{e, \tilde{N}_K} < \lambda_{e, \tilde{N}_{K+1}}$ ($R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$).

(ii) Using definition of L from equation (3), we first show below that

$$L_K < L_{K+1} \Leftrightarrow \sum_{i,j \geq 0} \sum_{i > j} (i-j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0, \quad (11)$$

(which also provides an alternative approach to prove Theorem 1/(ii)):

$$\begin{aligned} L_{K+1} > L_K & \Leftrightarrow \frac{\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)}{\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)} > \frac{\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n)}{\sum_{i=0}^{\infty} \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n)} \Leftrightarrow \frac{\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n)}{\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{K+1}(n)} > \frac{\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n)}{\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_K(n)}, \\ & \Leftrightarrow \left[\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \right] \left[\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_K(n) \right] > \left[\sum_{i=0}^{\infty} i \rho^i \prod_{n=0}^{i-1} \bar{F}_K(n) \right] \left[\sum_{j=0}^{\infty} \rho^j \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right], \\ & \Leftrightarrow \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i \rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) > \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i \rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n), \\ & \Leftrightarrow \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} i \rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right) > 0, \\ & \Leftrightarrow \sum_{i,j > 0} \sum_{i \neq j} i \rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right) > 0, \\ & \Leftrightarrow \sum_{i,j > 0} \sum_{i > j} i \rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right) \\ & \quad + \sum_{i,j > 0} \sum_{i < j} i \rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right) > 0, \\ & \Leftrightarrow \sum_{i,j > 0} \sum_{i > j} i \rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{i-1} \bar{F}_K(n) \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \right) \\ & \quad + \sum_{i,j > 0} \sum_{i > j} j \rho^{i+j} \left(\prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{i-1} \bar{F}_K(n) - \prod_{n=0}^{j-1} \bar{F}_K(n) \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \right) > 0. \end{aligned}$$

Regrouping again, gives

$$\sum_{i,j > 0} \sum_{i > j} \left((i-j) \rho^{i+j} \prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) + (j-i) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{i-1} \bar{F}_K(n) \right) > 0,$$

$$\begin{aligned}
&\Leftrightarrow \sum_{i,j>0} \sum_{i>j} (i-j)\rho^{i+j} \left(\prod_{n=0}^{i-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) - \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{i-1} \bar{F}_K(n) \right) > 0, \\
&\Leftrightarrow \sum_{i,j>0} \sum_{i>j} (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0.
\end{aligned}$$

Since, a_{K_n} is the largest value on the support of \bar{F}_K , we have $\bar{F}_{K+1}(i-1) = \bar{F}_K(i-1) = 0$ for $i \in \{a_{K_n} + 1, a_{K_n} + 2, \dots\}$. Hence, those indices can be dropped, which proves inequality (11):

$$L_K < L_{K+1} \Leftrightarrow \sum_{a_{K_n} \geq i} \sum_{j \geq 0} (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0.$$

$$\begin{aligned}
\text{Let us define } A_{K+1}(i, j) &\triangleq (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \prod_{n=j}^{i-1} \bar{F}_{K+1}(n); \\
A_K(i, j) &\triangleq (i-j)\rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \prod_{n=j}^{i-1} \bar{F}_K(n).
\end{aligned} \tag{12}$$

Then (11) reduces to $\sum_{a_{K_n} \geq i} \sum_{j \geq 0} [A_{K+1}(i, j) - A_K(i, j)] > 0$.

Similar to the approach used in the proof of part (i), our strategy is to form a partition of (i, j) based on the sign of $A_{K+1}(i, j) - A_K(i, j)$. The underlying space is the 2-dimensional set $\{(i, j) : a_{K_n} \geq i > j \geq 0\}$. Since $A_{K+1}(i, j) > A_K(i, j)$ if and only if $\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) > \prod_{n=j}^{i-1} \bar{F}_K(n)$, we shall seek a partition over $\{(i, j) : a_{K_n} \geq i > j \geq 0\}$ based on the sign of $\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n)$ instead.

Define $\mathcal{G}_1 \triangleq \{a_{K_n}\} \times \{0, 1, 2, \dots, a_{K_n} - 1\}$; $\mathcal{G}_2 \triangleq \{a_{K_1} + 1, a_{K_1} + 2, \dots, a_{K_n} - 2, a_{K_n} - 1\} \times \{0, 1, 2, \dots, a_{K_1}\}$; and $\mathcal{G}_3 \triangleq \{(i, j) : a_{K_n} \geq i > j \geq 0\} - \{\mathcal{G}_1 \cup \mathcal{G}_2\}$, i.e., \mathcal{G}_3 contains all the elements that are not in \mathcal{G}_1 or \mathcal{G}_2 .

From (8) and (9), we can verify that

$$\begin{aligned}
\forall (i, j) \in \mathcal{G}_1 : & \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) < \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i, j) - A_K(i, j) < 0; \\
\forall (i, j) \in \mathcal{G}_2 : & \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) > \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i, j) - A_K(i, j) > 0; \\
\forall (i, j) \in \mathcal{G}_3 : & \prod_{n=j}^{i-1} \bar{F}_{K+1}(n) = \prod_{n=j}^{i-1} \bar{F}_K(n) \Rightarrow A_{K+1}(i, j) - A_K(i, j) = 0.
\end{aligned}$$

Since \mathcal{G}_3 contains all (i, j) where $A_{K+1}(i, j) - A_K(i, j) = 0$, it suffices to just show that $\sum_{(i,j) \in \mathcal{G}_1 \cup \mathcal{G}_2} [A_{K+1}(i, j) - A_K(i, j)] > 0$ as a goal. Also since $A_{K+1}(i, j) - A_K(i, j) > 0, \forall (i, j) \in \mathcal{G}_2$, the inequality will hold if there exists a subset $\mathcal{G}_{2'} \subseteq \mathcal{G}_2$ such that

$$\sum_{(i,j) \in \mathcal{G}_1 \cup \mathcal{G}_{2'}} [A_{K+1}(i, j) - A_K(i, j)] > 0. \tag{13}$$

We shall prove that the sufficient condition on ρ stated in the lemma guarantees for inequality (13) to hold. To do that, we need to consider the elements of \mathcal{G}_1 and \mathcal{G}_2 in greater detail.

From the construction of the partition above, we have $|\mathcal{G}_1| = a_{K_n}$, i.e., there are a_{K_n} pairs of (i, j) in \mathcal{G}_1 , represented by $\{(a_{K_n}, a_{K_n} - 1), (a_{K_n}, a_{K_n} - 2), (a_{K_n}, a_{K_n} - 3), \dots, (a_{K_n}, 1), (a_{K_n}, 0)\}$. On the other hand, $|\mathcal{G}_2| = (a_{K_n} - a_{K_1} - 1)(a_{K_1} + 1)$. Treating $a_{K_n} - a_{K_1} - 1$ and $a_{K_1} + 1$ as base and height of a rectangular and

Order l	\mathcal{G}_1 contains:	$\mathcal{G}_{2'}$ contains:	Order l	\mathcal{G}_1 contains:	$\mathcal{G}_{2'}$ contains:
$l=1$	$(a_{K_n}, a_{K_n} - 1)$	$(a_{K_1} + 1, a_{K_1})$	$l=1$	$(a_{K_n}, a_{K_n} - 1)$	$(a_{K_1} + 1, a_{K_1})$
$l=2$	$(a_{K_n}, a_{K_n} - 2)$	$(a_{K_1} + 1, a_{K_1} - 1)$	$l=2$	$(a_{K_n}, a_{K_n} - 2)$	$(a_{K_1} + 1, a_{K_1} - 1)$
$l=3$	$(a_{K_n}, a_{K_n} - 3)$	$(a_{K_1} + 1, a_{K_1} - 2)$	$l=3$	$(a_{K_n}, a_{K_n} - 3)$	$(a_{K_1} + 1, a_{K_1} - 2)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$l=a_{K_1}-1$	$(a_{K_n}, a_{K_n} - a_{K_1} + 1)$	$(a_{K_1} + 1, 2)$	$l=a_{K_1}-1$	$(a_{K_n}, a_{K_n} - a_{K_1} + 1)$	$(a_{K_1} + 1, 2)$
$l=a_{K_1}$	$(a_{K_n}, a_{K_n} - a_{K_1})$	$(a_{K_1} + 1, 1)$	$l=a_{K_1}$	$(a_{K_n}, a_{K_n} - a_{K_1})$	$(a_{K_1} + 1, 1)$
$l=a_{K_1}+1$	$(a_{K_n}, a_{K_n} - a_{K_1} - 1)$	$(a_{K_1} + 1, 0)$	$l=a_{K_1}+1$	$(a_{K_n}, a_{K_n} - a_{K_1} - 1)$	$(a_{K_1} + 2, 1)$
$l=a_{K_1}+2$	$(a_{K_n}, a_{K_n} - a_{K_1} - 2)$	$(a_{K_1} + 2, 0)$	$l=a_{K_1}+2$	$(a_{K_n}, a_{K_n} - a_{K_1} - 2)$	$(a_{K_1} + 3, 1)$
$l=a_{K_1}+3$	$(a_{K_n}, a_{K_n} - a_{K_1} - 3)$	$(a_{K_1} + 3, 0)$	$l=a_{K_1}+3$	$(a_{K_n}, a_{K_n} - a_{K_1} - 3)$	$(a_{K_1} + 4, 1)$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$l=a_{K_n}-2$	$(a_{K_n}, 2)$	$(a_{K_n} - 2, 0)$	$l=a_{K_n}-2$	$(a_{K_n}, 2)$	$(a_{K_n} - 1, 1)$
$l=a_{K_n}-1$	$(a_{K_n}, 1)$	$(a_{K_n} - 1, 0)$	$l=a_{K_n}-1$	$(a_{K_n}, 1)$	
$l=a_{K_n}$	$(a_{K_n}, 0)$				

Table 3 \mathcal{G}_1 and $\mathcal{G}_{2'}$ used in the proof of Lemma 3/(ii) **Table 4** \mathcal{G}_1 and $\mathcal{G}_{2'}$ used in the proof of Lemma 3/(iii)

using the fact that a rectangular shape of fixed perimeter (a_{K_n}) contains less area ($|\mathcal{G}_2|$) when the shape is more asymmetric, we can show that $|\mathcal{G}_2| \geq a_{K_n} - 1$, with equality holds only when $a_{K_1} = a_{K_n} - 2$ (or $a_{K_1} = 0$ which is not possible).

The $a_{K_n} - 1$ pairs of (i, j) that are guaranteed to reside in \mathcal{G}_2 can be parametrized as $\{(a_{K_1} + 1, a_{K_1}), (a_{K_1} + 1, a_{K_1} - 1), (a_{K_1} + 1, a_{K_1} - 2), \dots, (a_{K_1} + 1, 1), (a_{K_1} + 1, 0), (a_{K_1} + 2, 0), (a_{K_1} + 3, 0), \dots, (a_{K_n} - 2, 0), (a_{K_n} - 1, 0)\}$. We therefore define this set to be $\mathcal{G}_{2'}$. Note that $|\mathcal{G}_1| = a_{K_n}$ and $|\mathcal{G}_{2'}| = a_{K_n} - 1$. We now order elements of \mathcal{G}_1 and $\mathcal{G}_{2'}$ in a specific way displayed in Table 3 (each element in either group is itself an (i, j) pair).

We denote $\mathcal{G}_1^l : l \in \{1, 2, \dots, a_{K_n}\}$ and $\mathcal{G}_{2'}^l : l \in \{1, 2, \dots, a_{K_n} - 1\}$ the l -th element in \mathcal{G}_1 and $\mathcal{G}_{2'}$, respectively, according to the order specified in Table 3. Furthermore, for each \mathcal{G}_1^l and each $\mathcal{G}_{2'}^l$, we specific its Cartesian coordinates by subscript i and j , i.e., $\mathcal{G}_1^l = (\{\mathcal{G}_1^l\}_i, \{\mathcal{G}_1^l\}_j)$ and $\mathcal{G}_{2'}^l = (\{\mathcal{G}_{2'}^l\}_i, \{\mathcal{G}_{2'}^l\}_j)$. For example, $\{\mathcal{G}_1^{a_{K_n}}\}_i = a_{K_n}$ and $\{\mathcal{G}_{2'}^{a_{K_n}}\}_j = 1$. We note that $\forall l \in \{1, 2, \dots, a_{K_n} - 1\}$,

$$\{\mathcal{G}_1^l\}_i = a_{K_n} = (a_{K_n} - 1) + 1 \geq \{\mathcal{G}_{2'}^l\}_i + 1 > \{\mathcal{G}_{2'}^l\}_i \quad (14)$$

$$\{\mathcal{G}_1^l\}_i - \{\mathcal{G}_1^l\}_j = l = \{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j \quad (15)$$

$$\{\mathcal{G}_1^l\}_i + \{\mathcal{G}_1^l\}_j = a_{K_n} + \{\mathcal{G}_1^l\}_j = 2a_{K_n} - l > 2(a_{K_n} - 1) - l \geq 2\{\mathcal{G}_{2'}^l\}_i - l \geq \{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j \quad (16)$$

Recall from (13), our goal is to find a sufficient condition such that the summation of $A_{K+1}(i, j) - A_K(i, j)$ over all (i, j) in $\mathcal{G}_1 \cup \mathcal{G}_{2'}$ is positive. We describe all the elements of $\mathcal{G}_1 \cup \mathcal{G}_{2'}$ by considering the first $(a_{K_n} - 2)$ rows of \mathcal{G}_1^l and $\mathcal{G}_{2'}^l$ in Table 3 plus the last three elements at the $(a_{K_n} - 1)$ -th and the a_{K_n} -th rows of the table (namely $\mathcal{G}_1^{a_{K_n}-1}$, $\mathcal{G}_{2'}^{a_{K_n}-1}$ and $\mathcal{G}_1^{a_{K_n}}$) from Table 3. Therefore, one set of sufficient conditions for (13) to hold is (a) $\forall l \in \{1, 2, \dots, a_{K_n} - 2\}$, $\sum_{(i,j) \in \{\mathcal{G}_1^l, \mathcal{G}_{2'}^l\}} [A_{K+1}(i, j) - A_K(i, j)] > 0$ and (b)

$$\sum_{(i,j) \in \{\mathcal{G}_1^{a_{K_n}-1}, \mathcal{G}_{2'}^{a_{K_n}-1}, \mathcal{G}_1^{a_{K_n}}\}} [A_{K+1}(i, j) - A_K(i, j)] > 0.$$

We first show that (a) is true for all ρ . $\forall l \in \{1, 2, \dots, a_{K_n} - 2\}$. Recall that $[A_{K+1}(i, j) - A_K(i, j)]$ evaluated at $(i, j) = \mathcal{G}_1^l$ is negative, and $[A_{K+1}(i, j) - A_K(i, j)]$ evaluated at $(i, j) = \mathcal{G}_{2'}^l$ is positive. It is thus equivalent to show that $[A_K(i, j) - A_{K+1}(i, j)] \Big|_{(i,j)=\mathcal{G}_1^l} < [A_{K+1}(i, j) - A_K(i, j)] \Big|_{(i,j)=\mathcal{G}_{2'}^l}$.

Denote $d = \min\{f_K(a_{K_1}), f_K(a_{K_n})\} > 0$, we have from (12) that

$$\begin{aligned}
& \left| [A_K(i, j) - A_{K+1}(i, j)] \right|_{(i,j)=\mathcal{G}_1^l} \\
&= (\{\mathcal{G}_1^l\}_i - \{\mathcal{G}_1^l\}_j) \rho^{\{\mathcal{G}_1^l\}_i + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{\{\mathcal{G}_1^l\}_i-1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{\{\mathcal{G}_1^l\}_i-1} \bar{F}_{K+1}(n) \right) \\
&= (a_{K_n} - \{\mathcal{G}_1^l\}_j) \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n}-1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n}-1} \bar{F}_{K+1}(n) \right) \\
&= l \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n}-1} \bar{F}_K(n) - \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n}-1} \bar{F}_{K+1}(n) \right) \\
&< l \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_K(n) \left(d \cdot \prod_{n=\{\mathcal{G}_1^l\}_j}^{a_{K_n}-2} \bar{F}_K(n) \right) \\
&= l \cdot d \cdot \rho^{a_{K_n} + \{\mathcal{G}_1^l\}_j} \prod_{n=0}^{\{\mathcal{G}_1^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \\
&< (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \\
&\quad \text{because } l = \{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j \text{ see (15), } \rho < 1 \text{ and } a_{K_n} + \{\mathcal{G}_1^l\}_j > \{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j \text{ see (16)} \\
&\leq (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \\
&\leq (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \cdot d \cdot \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=a_{K_1}+1}^{\{\mathcal{G}_{2'}^l\}_i-1} \bar{F}_K(n) \text{ because } a_{K_n} - 2 \geq \{\mathcal{G}_{2'}^l\}_i - 1 \text{ see (14)} \\
&= (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_K(n) \left(d \cdot \prod_{n=a_{K_1}+1}^{\{\mathcal{G}_{2'}^l\}_i-1} \bar{F}_K(n) \right) \\
&\quad \text{because } \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_K(n) = 1 \\
&= (\{\mathcal{G}_{2'}^l\}_i - \{\mathcal{G}_{2'}^l\}_j) \rho^{\{\mathcal{G}_{2'}^l\}_i + \{\mathcal{G}_{2'}^l\}_j} \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{\{\mathcal{G}_{2'}^l\}_j-1} \bar{F}_K(n) \left(\prod_{n=\{\mathcal{G}_{2'}^l\}_j}^{\{\mathcal{G}_{2'}^l\}_i-1} \bar{F}_{K+1}(n) - \prod_{n=\{\mathcal{G}_{2'}^l\}_j}^{\{\mathcal{G}_{2'}^l\}_i-1} \bar{F}_K(n) \right) \\
&\quad \text{because } \prod_{n=\{\mathcal{G}_{2'}^l\}_j}^{a_{K_1}-1} \bar{F}_{K+1}(n) = \prod_{n=\{\mathcal{G}_{2'}^l\}_j}^{a_{K_1}-1} \bar{F}_K(n) = 1, \bar{F}_{K+1}(a_{K_1}) - \bar{F}_K(a_{K_1}) = d \\
&\quad \text{and } \bar{F}_{K+1}(x) = \bar{F}_K(x), \forall x \in \{a_{K_1} + 1, \dots, \{\mathcal{G}_{2'}^l\}_i - 1\} \\
&= [A_{K+1}(i, j) - A_K(i, j)] \Big|_{(i,j)=\mathcal{G}_{2'}^l}, \text{ as required.}
\end{aligned}$$

Next, for (b) to hold, i.e., $\sum_{(i,j) \in \{\mathcal{G}_1^{a_{K_n}-1}, \mathcal{G}_{2'}^{a_{K_n}}, \mathcal{G}_1^{a_{K_n}}\}} [A_{K+1}(i, j) - A_K(i, j)] > 0$, we have equivalently:

$$\begin{aligned}
& [A_K(i, j) - A_{K+1}(i, j)] \Big|_{(i,j)=\mathcal{G}_1^{a_{K_n}-1}} + [A_K(i, j) - A_{K+1}(i, j)] \Big|_{(i,j)=\mathcal{G}_{2'}^{a_{K_n}}} < [A_{K+1}(i, j) - A_K(i, j)] \Big|_{(i,j)=\mathcal{G}_{2'}^{a_{K_n}-1}} \Leftrightarrow \\
& [A_K(i, j) - A_{K+1}(i, j)] \Big|_{(i,j)=(a_{K_n}, 1)} + [A_K(i, j) - A_{K+1}(i, j)] \Big|_{(i,j)=(a_{K_n}, 0)} < [A_{K+1}(i, j) - A_K(i, j)] \Big|_{(i,j)=(a_{K_n}-1, 0)} \quad (17)
\end{aligned}$$

Note that (with any empty product being equal to $= 1$)

$$\begin{aligned}
& \left| [A_K(i, j) - A_{K+1}(i, j)] \right|_{(i, j) = (a_{K_n}, 1)} = (a_{K_n} - 1) \rho^{a_{K_n} + 1} \cdot \left(\prod_{n=1}^{a_{K_n}-1} \bar{F}_{K+1}(n) - \prod_{n=1}^{a_{K_n}-1} \bar{F}_K(n) \right) \\
& < (a_{K_n} - 1) \rho^{a_{K_n} + 1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\
& \left| [A_K(i, j) - A_{K+1}(i, j)] \right|_{(i, j) = (a_{K_n}, 0)} = a_{K_n} \rho^{a_{K_n}} \cdot \left(\prod_{n=0}^{a_{K_n}-1} \bar{F}_{K+1}(n) - \prod_{n=0}^{a_{K_n}-1} \bar{F}_K(n) \right) \\
& < a_{K_n} \rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\
& \left| [A_{K+1}(i, j) - A_K(i, j)] \right|_{(i, j) = (a_{K_n}-1, 0)} = (a_{K_n} - 1) \rho^{a_{K_n}-1} \cdot \left(\prod_{n=0}^{a_{K_n}-2} \bar{F}_{K+1}(n) - \prod_{n=0}^{a_{K_n}-2} \bar{F}_K(n) \right) \\
& = (a_{K_n} - 1) \rho^{a_{K_n}-1} \cdot \left(\prod_{n=a_{K_1}}^{a_{K_n}-2} \bar{F}_{K+1}(n) - \prod_{n=a_{K_1}}^{a_{K_n}-2} \bar{F}_K(n) \right) = (a_{K_n} - 1) \rho^{a_{K_n}-1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n).
\end{aligned}$$

Therefore, it is sufficient for (17) to hold if

$$\begin{aligned}
& (a_{K_n} - 1) \rho^{a_{K_n} + 1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) + a_{K_n} \rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) \leq (a_{K_n} - 1) \rho^{a_{K_n}-1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) \\
& \Leftrightarrow (a_{K_n} - 1) \rho^{a_{K_n} + 1} + a_{K_n} \rho^{a_{K_n}} \leq (a_{K_n} - 1) \rho^{a_{K_n}-1} \Leftrightarrow \rho^2 + \frac{a_{K_n}}{a_{K_n} - 1} \rho \leq 1.
\end{aligned} \tag{18}$$

Solving quadratic equation (18) gives the condition

$$\frac{1}{2} \left(-\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right) \leq \rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right),$$

where it is clear that $\frac{1}{2} \left(-\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right) < 0$ and $0 < \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right) < 1$.

Since $\rho \in (0, 1)$, we conclude that, when $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right)$, (18) (17), (13) and (11) all hold and thus it is a sufficient condition for $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$. This completes the proof of part (ii).

(iii) Using the definition of W from equation (6), and comparing the structure of equation (3) to that of (6), it can be shown via a similar approach used in the proof of inequality (11) that

$$W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}} \Leftrightarrow \sum_{i, j \geq 1: i > j} (i - j) \rho^{i+j} \prod_{n=0}^{j-1} \bar{F}_{K+1}(n) \prod_{n=0}^{j-1} \bar{F}_K(n) \left(\prod_{n=j}^{i-1} \bar{F}_{K+1}(n) - \prod_{n=j}^{i-1} \bar{F}_K(n) \right) > 0, \tag{19}$$

which also provides an alternative approach to prove Theorem 1/(iii). Using the same definition of $A_K(i, j)$ and $A_{K+1}(i, j)$ from part (ii), we have $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}} \Leftrightarrow \sum_{a_{K_n} \geq i > j \geq 1} [A_{K+1}(i, j) - A_K(i, j)] > 0$. The rest of the proof is then almost identical to the proof of part (ii) except now i, j cannot take on 0. Define $\mathcal{G}_1 \triangleq \{a_{K_n}\} \times \{1, 2, \dots, a_{K_n} - 1\}$, $\mathcal{G}_2 \triangleq \{a_{K_1} + 1, a_{K_1} + 2, \dots, a_{K_n} - 2, a_{K_n} - 1\} \times \{1, 2, \dots, a_{K_1}\}$, and $\mathcal{G}_3 \triangleq \{(i, j) : a_{K_n} \geq i > j \geq 1\} - \{\mathcal{G}_1 \cup \mathcal{G}_2\}$.

There are now at least $a_{K_n} - 2$ elements in the set \mathcal{G}_2 which defines the subset $\mathcal{G}_{2'}$. The elements in \mathcal{G}_1 and $\mathcal{G}_{2'}$ are ordered in a similar fashion as before and displayed in Table 4. A sufficient condition for

$W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ from (19) is that $\sum_{(i,j) \in \mathcal{G}_1 \cup \mathcal{G}_{2'}} [A_{K+1}(i,j) - A_K(i,j)] > 0$. It can be shown that conditions (14)-(16) still hold, and thus $\forall l \in \{1, 2, \dots, a_{K_n} - 3\}$ and for all ρ , $\sum_{(i,j) \in \{\mathcal{G}_1^l, \mathcal{G}_{2'}^l\}} [A_{K+1}(i,j) - A_K(i,j)] > 0$. Therefore, a sufficient condition for $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ is that

$$\sum_{(i,j) \in \{\mathcal{G}_1^{a_{K_n}-2}, \mathcal{G}_{2'}^{a_{K_n}-2}, \mathcal{G}_1^{a_{K_n}-1}\}} [A_{K+1}(i,j) - A_K(i,j)] > 0. \quad (20)$$

$$\begin{aligned} \text{Since } [A_K(i,j) - A_{K+1}(i,j)] \Big|_{(i,j)=(a_{K_n}, 2)} &< (a_{K_n} - 2)\rho^{a_{K_n}+2} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\ [A_K(i,j) - A_{K+1}(i,j)] \Big|_{(i,j)=(a_{K_n}, 1)} &< (a_{K_n} - 1)\rho^{a_{K_n}+1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \\ \text{and } [A_{K+1}(i,j) - A_K(i,j)] \Big|_{(i,j)=(a_{K_n}-1, 1)} &= (a_{K_n} - 2)\rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n), \end{aligned}$$

condition (20) will hold if

$$\begin{aligned} (a_{K_n} - 2)\rho^{a_{K_n}+2} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) + (a_{K_n} - 1)\rho^{a_{K_n}+1} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) &\leq (a_{K_n} - 2)\rho^{a_{K_n}} \cdot d \cdot \prod_{n=a_{K_1}+1}^{a_{K_n}-2} \bar{F}_K(n) \\ \Leftrightarrow (a_{K_n} - 2)\rho^{a_{K_n}+2} + (a_{K_n} - 1)\rho^{a_{K_n}+1} &\leq (a_{K_n} - 2)\rho^{a_{K_n}} \Leftrightarrow \rho^2 + \frac{a_{K_n} - 1}{a_{K_n} - 2}\rho \leq 1. \end{aligned}$$

The solution of the quadratic inequality on the set $\rho \in (0, 1)$ is $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}-1}{a_{K_n}-2}\right)^2 + 4} - \frac{a_{K_n}-1}{a_{K_n}-2} \right)$. \square

Proof of Theorem 2:

- (i) Result follows immediately from Lemma 3/(i) since $\lambda_{eff, \tilde{N}_K} < \lambda_{eff, \tilde{N}_{K+1}}$ ($R_{\tilde{N}_K} < R_{\tilde{N}_{K+1}}$) for all K .
- (ii) Recall from Lemma 3/(ii) that $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ if $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right)$. It can be easily verified that $\frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}}{a_{K_n}-1}\right)^2 + 4} - \frac{a_{K_n}}{a_{K_n}-1} \right)$ increases in a_{K_n} . Plugging in the smallest possible value of a_{K_n} which is 3, we get $\rho = 0.5$. Therefore, when $\rho \leq 0.5$, $L_{\tilde{N}_K} < L_{\tilde{N}_{K+1}}$ for all K (regardless of the distributions of $\{\tilde{N}_K\}_{K=0,1,2,\dots,T}$). Result thus follows. Note it is possible to derive stronger distribution-specific conditions.
- (iii) Recall from Lemma 3/(iii) that $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ if $\rho \leq \frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}-1}{a_{K_n}-2}\right)^2 + 4} - \frac{a_{K_n}-1}{a_{K_n}-2} \right)$. It can be verified that $\frac{1}{2} \left(\sqrt{\left(\frac{a_{K_n}-1}{a_{K_n}-2}\right)^2 + 4} - \frac{a_{K_n}-1}{a_{K_n}-2} \right)$ increases in a_{K_n} . Plugging in the smallest possible value of a_{K_n} which is 3, we get $\rho = 0.414$. Therefore, when $\rho \leq 0.414$, $W_{\tilde{N}_K} < W_{\tilde{N}_{K+1}}$ for all K . Result thus follows. Again, it is possible to derive stronger distribution-specific conditions. \square

Proof of Proposition 1:

Consider the random variable $\tilde{N}_T \in \{\lfloor \mathbb{E}(\tilde{N}) \rfloor, \lceil \mathbb{E}(\tilde{N}) \rceil\}$ such that $E(\tilde{N}_T) = E(\tilde{N})$. By Theorem 2, we have $R_{\tilde{N}} \leq R_{\tilde{N}_T}$ for all ρ , and $L_{\tilde{N}} \leq L_{\tilde{N}_T}$, $W_{\tilde{N}} \leq W_{\tilde{N}_T}$ for small ρ . On the other hand, since $\mathbb{E}(\tilde{N}) \leq N$ and N is an integer, we must have $\lceil \mathbb{E}(\tilde{N}) \rceil \leq N$. It follows that $\tilde{N}_T \leq_{st} N$ so by Theorem 1, we have $R_{\tilde{N}_T} \leq R_N$ for all ρ , and $L_{\tilde{N}_T} \leq L_N$, $W_{\tilde{N}_T} \leq W_N$ for small ρ . Result thus follows. \square

Proof of Lemma 4

From the properties of a floor function, we know that $\frac{\mu(v-p)}{c} - 1 < N = \lfloor \frac{\mu(v-p)}{c} \rfloor \leq \frac{\mu(v-p)}{c}$, and $\frac{\tilde{\mu}(v-p)}{c} - 1 <$

$\tilde{N} = \lfloor \frac{\tilde{\mu}(v-p)}{c} \rfloor \leq \frac{\tilde{\mu}(v-p)}{c}$ for every realization $\Rightarrow \frac{\mathbb{E}(\tilde{\mu})(v-p)}{c} - 1 < \mathbb{E}(\tilde{N}) \leq \frac{\mathbb{E}(\tilde{\mu})(v-p)}{c}$. (i) When $\mathbb{E}(\tilde{\mu}) \leq \mu - \frac{c}{v-p}$, $\mathbb{E}(\tilde{N}) \leq \frac{\mu(v-p)}{c} - 1 < N$. (ii) When $\mathbb{E}(\tilde{\mu}) \geq \mu + \frac{c}{v-p}$, $\mathbb{E}(\tilde{N}) > \frac{\mu(v-p)}{c} \geq N$. (iii) When $|\mathbb{E}(\tilde{\mu}) - \mu| < x \cdot \frac{c}{v-p}$ for some $x \in (0, 1]$, we will discuss three sub-cases: (a) If $\mathbb{E}(\mu) = \mu$, then $\frac{\mu(v-p)}{c} - 1 < \mathbb{E}(\tilde{N}) \leq \frac{\mu(v-p)}{c}$ and $|\mathbb{E}(\tilde{N}) - N| < 1$; (b) If $\mu < \mathbb{E}(\mu) < \mu + x \cdot \frac{c}{v-p}$, then $\frac{\mu(v-p)}{c} - 1 < \frac{\mathbb{E}(\mu)(v-p)}{c} - 1 < \mathbb{E}(\tilde{N}) \leq \frac{\mathbb{E}(\mu)(v-p)}{c} < \frac{\mu(v-p)}{c} + x$ and $|\mathbb{E}(\tilde{N}) - N| < 1 + x$; (c) If $\mu - x \cdot \frac{c}{v-p} < \mathbb{E}(\mu) < \mu$, then $\frac{\mu(v-p)}{c} - x - 1 < \frac{\mathbb{E}(\mu)(v-p)}{c} - 1 < \mathbb{E}(\tilde{N}) \leq \frac{\mathbb{E}(\mu)(v-p)}{c} < \frac{\mu(v-p)}{c}$ and $|\mathbb{E}(\tilde{N}) - N| < 1 + x$. Combining the three sub-cases, it is easy to see that $|\mathbb{E}(\tilde{N}) - N| < 1 + x$. \square

References

- Akşin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.
- Allon, G., A. Bassamboo, Q. Yu. 2014. The impact of delay announcements on consumers: an empirical study. *Kellogg Working Paper*.
- Armony, M., C. Maglaras. 2004a. Contact centers with a call-back option and real-time delay information. *Operations Research* **52**(4) 527–545.
- Armony, M., C. Maglaras. 2004b. On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Operations Research* **52**(2) 271–292.
- Berger, A., W. Whitt. 1992. Comparisons of multi-server queues with finite waiting rooms. *Communications in Statistics – Stochastic Models* **8**(4) 719–732.
- Besbes, O., B. Dooley, N. Gans. 2011. Dynamic service control of a queue with congestion-sensitive customers. *Working Paper*.
- Besbes, O., C. Maglaras. 2009. Revenue optimization for a make-to-order queue in an uncertain market environment. *Operations Research* **57**(6) 1438–1450.
- Bhaskaran, B. 1986. Almost sure comparison of birth and death processes with application to $m/m/s$ queueing systems. *Queueing Systems* **1**(1) 103–127.
- Debo, L. G., S. Veeraraghavan. 2014. Equilibrium in queues under unknown service rates and service value. *Operations Research* **62**(2) 38–57.
- Economou, A., S. Kanta. 2008. Optimal balking strategies and pricing for the single server markovian queue with compartmented waiting space. *Queueing Systems* **59**(3) 237–269.
- Edelson, N. M., D. K. Hildebrand. 1975. Congestion tolls for poisson queueing processes. *Econometrica: Journal of the Econometric Society* 81–92.
- Guo, P., W. Sun, Y. Wang. 2011. Equilibrium and optimal strategies to join a queue with partial information on service times. *European Journal of Operational Research* **214**(2) 284–297.
- Guo, P., P. Zipkin. 2007. Analysis and comparison of queues with different levels of delay information. *Management Science* **53**(6) 962–970.

- Guo, P., P. Zipkin. 2009. The effects of the availability of waiting-time information on a balking queue. *European Journal of Operational Research* **198**(1) 199–209.
- Hassin, R. 1986. Consumer information in markets with random product quality: The case of queues and balking. *Econometrica* **54**(5) 1185–1195.
- Hassin, R. 2007. Information and uncertainty in a queuing system. *Probability in the Engineering and Informational Sciences* **21**(3) 361.
- Hassin, R., M. Haviv. 2003. *To queue or not to queue: Equilibrium behavior in queueing systems*, vol. 59. Springer International Series in Operations Research & Management Science.
- Haviv, M., R. Randhawa. 2014. Pricing in queues without demand information. *Manufacturing & Service Operations Management* **16**(3) 401–411.
- Huang, T., G. Allon, A. Bassamboo. 2013. Bounded rationality in service systems. *Manufacturing & Service Operations Management* **15**(2) 263–279.
- Huang, T., Y. Chen. 2014. Service systems with experience-based anecdotal reasoning customers. *Production and Operations Management* .
- Jouini, O., Z. Aksin, Y. Dallery. 2011. Call centers with delay information: Models and insights. *Manufacturing & Service Operations Management* **13**(4) 534–548.
- Larsen, C. 1998. Investigating sensitivity and the impact of information on pricing decisions in an M/M/1/queueing model. *International journal of production economics* **56** 365–377.
- Lu, Y., A. Musalem, M. Olivares, A. Schilkrut. 2013. Measuring the effect of queues on customer purchases. *Management Science* **59**(8) 1743–1763.
- Luce, R. 1959. *Individual Choice Behavior*. Wiley, New York.
- Müller, A., D. Stoyan. 2002. *Comparison methods for stochastic models and risks*, vol. 389. Wiley Series in Probability and Statistics.
- Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* **37**(1) 15–24.
- Quirk, J. P., R. Saposnik. 1962. Admissibility and measurable utility functions. *The Review of Economic Studies* 140–146.
- Rothschild, M., J. E. Stiglitz. 1970. Increasing risk: I. a definition. *Journal of Economic Theory* **2**(3) 225–243.
- Su, X. 2008. Bounded rationality in newsvendor models. *Manufacturing & Service Operations Management* **10**(4) 566–589.
- Whitt, W. 1999. Improving service by informing customers about anticipated delays. *Management science* **45**(2) 192–207.
- Xu, S., L. Gao, J. Ou. 2007. Service performance analysis and improvement for a ticket queue with balking customers. *Management Science* **53**(6) 971–990.